



2

SOCIETATE & CUNOAȘTERE

De la procesarea de imagini către vederea artificială. Progrese cognitive

Mihaela Costin

(coordonator)



UNIA EUROPEANĂ



GUVERNUL ROMÂNIEI
MINISTERUL ÎNĂLȚĂRII
PROTECȚIEI SOCIALE ȘI
PERSONELOR VÂRSTNICE
AMPOSDRU



Fondul Social European
POSDRU 2007-2013



Instrumente Structurale
2007-2013



MINISTERUL
EDUCAȚIEI
NAȚIONALE

OLPOSDRU



ACADEMIA ROMÂNĂ
Filiala Iași

INSTITUTUL EUROPEAN

Mihaela COSTIN (coordonator) este cercetător științific III în cadrul Academiei Române, filiala Iași, a susținut doctoratul în inginerie în cadrul Universității Tehnice „Gh. Asachi”, Iași, cu teza „Contribuții la dezvoltarea sistemelor de decizie și aplicații” în anul 2005, a absolvit Facultatea de Electronică și Telecomunicații, UTI, în 1982 și a urmat studii post-universitare în economie, psihologie și științele educației, în cadrul Universității „Al. I. Cuza”. Cadru didactic asociat la Facultatea de Informatică U.A.I.C. și la Facultatea de Automatică și Calculatoare, U.T.I. (1996-2004), activitățile sale didactice și de cercetare științifică din cadrul colectivului de procesare de imagini al Institutului de Informatică Teoretică al Academiei Române (din 1989 până în prezent http://www.iit.tuiasi.ro/personal/m_costin.html) s-au concretizat în lucrări, cărți, contracte științifice și proiecte aplicative ale domeniilor de inteligență artificială, științe cognitive, procesări 1D, 2D. Bursier post-doctoral în proiectul POSDRU 56815 „Societatea bazată pe cunoaștere, cercetări, dezbateri, perspective”, a inițiat o colaborare științifică cu IRRIT, Universitatea Paul Sabatier, Toulouse, care a condus la editarea acestui volum.

© 2013 Academia Română – Filiala Iași

Iași, str. Grigore Ghica Vodă nr. 13, O. P. 1, C.P. 161
www.euroinst.ro; euroedit@hotmail.com

Descrierea CIP a Bibliotecii Naționale a României
De la procesarea de imagini către vederea artificială. Progrese
cognitive / ed., pref.: Mihaela Costin. - Iași : Institutul European, 2013
Bibliogr.

004
D28

MIHAELA COSTIN
(coordonator)

De la procesarea de imagini către vederea artificială Progrese cognitive

Cuvânt înainte de MIHAELA COSTIN

De la procesarea de imagini către vederea artificială Progrese cognitive

CUPRINS

Cuvânt-înainte (Mihaela Costin)	7
Aspecte cognitive în indexarea automată a conținutului audiovizual (Philippe Joly)	10
Tehnologii de asistare a persoanelor cu deficiențe vizuale (Christophe Jouffrais)	36
Dispozitive imagistice dezvoltate de Centrul de Excelență Secolul 21 Grupul Universității Shizuoka. O introducere în nanovision (Hidenori Mimura)	96
Stabilirea corespondențelor între pixeli în stereoviziunea binoculară (Alain Crouzil, Guillaume Gales, Sylvie Chambon)	114
Tehnici de reconstrucție 3D din imagini multiple (Florin Rotaru, Silviu Bejinariu, Simona Pescaru)	182
Metode de analiză și extragere a caracteristicilor esențiale din imaginile digitale (Adrian Ciobanu)	229
De la percepția vizuală către detecția automată a texturii (Mihaela Costin și Anca Ignat)	257
Percepție, reprezentare și limitele posibilităților de cunoaștere (Dan Gabriel Sîmbotin)	284
Index	295
Note despre autori	298

From image processing to artificial vision
Cognitive advances

SUMMARY

<i>Foreword</i> (Mihaela Costin)	7
Cognitive Issues in Automatic Audiovisual Content Indexing (Philippe Joly)	10
Assistive Technologies for the Visually Impaired Persons (Christophe Jouffrais)	36
Imaging Devices Developed by the 21st Century COE Group in Shizuoka University. An Introduction to Nanovision (Hidenori Mimura)	96
Pixel Matching in Binocular StereoVision (Alain Crouzil, Guillaume Gales, Sylvie Chambon)	114
3D Reconsruction Techniques from Multiple Images (Florin Rotaru, Silviu Bejinariu, Simona Pescaru)	182
Methods of Analysis and Extraction of Essential Digital Image Features (Adrian Ciobanu)	229
From visual perception to automatic detection of texture (Mihaela Costin and Anca Ignat)	257
Perception, Representation and Limits of Knowing (Dan Gabriel Sîmbotin)	284
Index	295
Notes abouts authors	298

Cuvânt-înainte

Motto: „Everyone takes the limits of his own vision for the limits of the world.”

Arthur Schopenhauer

Evoluând „de la procesarea de imagini către vederea artificială”... luând în calcul progresele cognitive realizate în ultimele decenii, cartea ne poartă logic printr-un număr de aplicații și teme de cercetare moderne, inedite, cu o deosebită aplicabilitate practică și didactică.

Subiectele abordate, deși variate, au un loc comun în termenul generos, relativ dificil traductibil, *vision*.

De la viziunea ideatică, la percepția vizuală, asistarea deficiențelor vizuale, măsurători nanometrice prin nanovision, la vederea artificială, sau computerizată, utilizarea diferită a termenilor în traducerea noțiunii *vision*, se datorează, de fiecare dată, unui alt context. Tehnologiile moderne de explorări funcționale permit descifrarea structurilor complexe ale cortexului și reușesc să decodifice modalitatea în care se realizează percepția vizuală și semantica gândirii asociate.

Gravitând în jurul acestui „nucleu terminologic”, *vision*, cartea este rezultatul colaborării cu cercetătorii din cadrul Institutului de Cercetări în Informatică din Toulouse, IRIT, Universitatea Paul Sabatier, realizată majoritar în timpul stagiului post-doctoral pe care l-am efectuat în cadrul proiectului POSDRU 56815, „Societatea bazată pe cunoaștere, cercetări, dezbateri, perspective”, dar și al oportunității de a cunoaște cercetători în domenii de vârf, în timpul conferințelor și al colaborărilor profesionale ocazionate de proiect, cum este domeniul nano-vision.

Capitolul care deschide această incursiune în zona aplicațiilor tehnologice conexe vederii umane și artificiale, „Aspecte cognitive în indexarea automată a conținutului audiovizual” (în forma sa originală sub denumirea „Cognitive Issues in Automatic Audiovisual Content Indexing”), este scris de profesorul Philippe Joly, specialist în domeniul structurării automate a conținutului audiovizual, coordonatorul unui proiect vast în IRIT și Universitatea Toulouse III, reunind patru grupuri de cercetare prin tema „Analiză și sinteză de semnal și imagine”, responsabil ERASMUS cu probleme de informatică al Universității Toulouse III.

Trecerea în revistă a metodelor de indexare a conținutului audiovizual al înregistrărilor și retrospectiva cercetărilor personale și de colectiv din ultima decadă (echipa SAMOVA – structurare, analiză și modelare a conținutului video și audio, din IRIT), ne conduc către aspectele cognitive explorate în regia artistică, în film și multimedia, pentru inducerea efectelor speciale prin asocieri imagini-sunet.

Christophe Jouffrais, director de cercetări al noului laborator LACII (Laboratoire Commun IRIT - IJA - Institut des Jeunes Aveugles de Toulouse), creat în toamna anului 2012 pentru a deservi necesitățile tinerilor nevăzători într-o societate din ce în ce mai tehnologizată, ne prezintă cercetările și proiectele inițiate pentru dezvoltarea unor noi „Tehnologii de asistare a persoanelor cu deficiențe vizuale” (engl. „Assistive Technologies for the Visually

Impaired Persons"). Realizările acestui laborator impresionează prin interactivitatea dispozitivelor elaborate, a mediului virtual simulat prin software dedicat identificării drumului optim de către persoana cu deficiențe vizuale, prin planificarea minuțioasă, asistată de calculator și memorarea căii de parcurs. Specialist în neuroscience, această cercetare științifică pe care o coordonează vine în întâmpinarea unui proiect ambițios, similar implantului cochlear, proteza vizuală.

Profesorul Hidenori Mimura, directorul Institutului de Cercetare în Electronică, Universitatea Shizuoka, Japonia, realizează o introducere în *nanovision*, prezentându-ne câteva „Dispozitive imagistice dezvoltate de Centrul de Excelență Secolul 21 – Grupul Universității Shizuoka”, (engl. „Imaging Devices Developed by the 21st Century COE Group in Shizuoka University, an Introduction to Nanovision”). Microscopul cu fluorescență care poate detecta structurile de ordin nanometric prin măsurarea timpului de descreștere a emisiei fluorescente este doar una dintre realizările extrem de interesante prezentate în acest capitol.

Alain Crouzil, Guillaume Gales, Sylvie Chambon analizează modalități noi de „Stabilire a corespondențelor între pixeli în stereoviziunea binoculară” (fr. "Mise en correspondance de pixels pour la stéréovision binoculaire", engl. „Pixel Matching in binocular StereoVision”). Alain Crouzil, conferențiar la Universitatea Paul Sabatier, este un dascăl entuziast și talentat care a dat acestui text o deosebită valoare didactică. Fiind descrisă în amănunt procedura stabilirii corespondențelor stânga-dreapta pentru interpretarea și reconstrucția imaginilor, metodele prezentate sunt foarte utile și în dezvoltarea unor noi proiecte de cercetare. Și acest capitol a fost elaborat în timpul stagiului post-doctoral la IRIT, UPS, Toulouse, Franța.

Pe aceeași temă se dezvoltă și capitolul scris de Florin Rotaru, Silviu Bejinariu, Simona Pescaru*, cercetători științifici în cadrul Academiei Române, Institutul de Informatică Teoretică, respectiv *, „vision scientist” în cadrul societății Robotic Vision Technologies. Abordând noi aspecte, se continuă tema precedentă, prin „Tehnici de reconstrucție 3D din imagini multiple” (engl. „3D Reconsruction Techniques from Multiple Images”). Cele două capitole se completează reciproc prin detalii tehnice cu suport fundamentat matematic.

Dacă primul capitol se referă la indexarea automată a conținutului unor înregistrări pe baza trăsăturilor extrase, o continuare logică o găsim în capitolul „Metode de analiză și extragere a caracteristicilor esențiale din imaginile digitale”, în care Adrian Ciobanu prezintă amănunțit modalitățile de extragere a unor trăsături esențiale, economice, din imagini și aparatul lor simplu de calcul, destinat regăsirii automate a imaginilor și clasificării acestora în baze masive de date.

„De la percepția vizuală către detecția automată a texturii” (engl. „From visual perception to automatic detection of texture”), autori Mihaela Costin (cercetător științific în cadrul Academiei Române) și Anca Ignat (cadru didactic în cadrul Universității „Al. I. Cuza”), este un capitol care analizează elementele neuro-fiziologice ce se pretează modelării, prin parcurgerea unor studii în domeniul percepției vizuale, focalizându-se pe textură. Dificil de realizat, testul perceptiv asupra texturii beneficiază mai mult de concluzii din teste conexe, decât directe. Tragem concluzii asupra percepției texturii din testele de detecție a orientării, viteza de observare a unui animal pe un anumit background, testele de identificare a luminozității sau de percepere a umbrelor etc. Testele realizate prin detectarea amplitudinii și direcționalității în texturi utilizând metoda Dual Tree Complex Wavelet Transform, prezintă o oarecare similaritate cu modul în care coloane de neuroni sesizează aceste detalii, în sistemul vizual uman și nu numai. Compararea mai multor tehnici laborioase de calcul al similarității în detecția texturii, pe imagini bidimensionale, utilizând baze extinse de imagini cu textură, ne dau doar un indiciu despre calea laborioasă de parcurs către ceea ce ne place să numim *vedere artificială*, „*artificial vision*”.

Ca un corolar în finalul acestei lucrări domnul profesor Dan Gabriel Sîmbotin vine să descrie puntea dintre cercetările în psihologie și științe cognitive către epistemologie și filozofie în „Percepție, reprezentare și limitele posibilităților de cunoaștere” (engl. „Perception, Representation and Limits of Knowing”).

Cartea este rezultatul direct al proiectului „Societatea Bazată pe Cunoaștere – cercetări, dezbateri, perspective”, cofinanțat de Uniunea Europeană și Guvernul României din Fondul Social European prin Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007–2013, POSDRU 56815, și a fost concepută, discutată, scrisă și tradusă cu ocazia stagiului pe care l-am efectuat la IRIT, UPS, Toulouse, Franța și al participării la 11th International Conference on Global Research and Education, 2012, InterAcademia august 2012, Budapesta.

Mulțumesc bunilor mei colegi Adrian Ciobanu, Silviu Bejinariu, Florin Rotaru, fizicienilor Cristina Niță și profesorului Dumitru Luca, care mi-au stat alături cu sfaturi profesionale, discuții, completări și asistarea traducerii pe parcursul editării acestei lucrări.

Mulțumiri deosebite pentru colega mea, Anca Ignat (cadru didactic la Facultatea de Informatică, UAIC), cu care, de câțiva ani buni, analizăm metodele de caracterizare a texturilor.

Alese gânduri de recunoștință celor care au făcut posibilă publicarea acestei cărți despre *vision*, îndrumătorilor din proiect, profesorilor Nicu Gavriluță, Cornelia Găspărel, Dan Gabriel Sîmbotin, Ovidiu Gherasim, acad. Teodor Dima și tuturor celor care au pus mult suflet și energie pentru buna desfășurare a proiectului, și care merită toată admirația noastră, doamnei Meda Gâlea, Irina Frasin, Margareta Munteanu, Giorgia Donceag, Magdalena Lazăr, domnului Codrin Dinu Vasiliu.

Cu plăcere gândurile mi se îndreaptă către cercetătorii din IRIT, Institutul de Cercetări în Informatică din Toulouse, centru similar Institutului de Informatică Teoretică al Academiei Române, care m-au primit cu o deosebită colegialitate și deschidere. Elaborarea acestor capitole a implicat un număr important de reuniuni de lucru, discuții, corecturi, concretizate în timp afectat din programul de lucru extrem de încărcat al unor profesori și cercetători renumiți cum sunt: Alain Crouzil, Philippe Joly, Christophe Jouffrais, dar și Regine Andre-Obrecht, Didier Dubois, Henri Prade, Simon Thorpe, Claudette Cayrol, Florence Dupin de Saint Cyr, Marie Christine Schiex și celor care m-au făcut să mă simt realmente colega lor de birou, în IRIT 1, etajul 3, sala 314, Yannick Chevalier și Philippe Balbiani.

Mulțumiri speciale profesorului Hidenori Mimura, care a afectat scrierii acestui capitol, un timp deosebit de prețios în domeniul *nanovision*, în cercetarea de înalt nivel pe care o coordonează ca director în Research Institute of Electronics, Shizuoka University, Japonia.

Cartea deschide un drum, *de la procesarea de imagini către vederea artificială prin studiul aspectelor cognitive* pentru toți cei entuziaști în a extinde sensurile și aplicațiile posibile ale domeniului *vision*.

Mihaela Costin
Iași, februarie, 2013.

Aspecte cognitive în indexarea automată a conținutului audiovizual

PHILIPPE JOLY¹

ABSTRACT. Audiovisual content indexing is the preliminary step to build a search engine to retrieve multimedia contents. The difficulty of this task is more important when no textual description accompanies the documents. The first scientific works made to address this topic were focusing on building discriminative digital signatures from basic features. But the need for indexing has grown with the amount of documents available to ordinary people and new strategies need deeper analysis. Some of the perspectives do take into account cognitive issues in these signatures or in the distance computation between them. In this chapter, we address several recent issues on scene classification for image retrieval, similarity estimation, character identification and temporal structure analysis.

KEYWORDS: *content indexing, scene, object, multimedia.*

1. Introducere

Indexarea conținutului audiovizual este doar un pas preliminar în construcția unui motor de căutare dedicat preluării unui conținut multimedia. Sarcina este evident mai dificilă atunci când nici o descriere textuală nu însoțește documentele. Primele lucrări științifice care au abordat această temă, s-au focalizat pe construirea de semnături digitale discriminative pornind de la trăsăturile de bază. Dar necesitatea indexării a crescut cu volumul de documente disponibile on-line, impunându-se descoperirea unor noi strategii. O posibilă perspectivă ia în considerație aspectele cognitive în atribuirea acestor semnături și în calculul distanțelor dintre ele. Vom aborda în cele ce urmează mai multe probleme recente cu privire la clasificarea scenelor pentru regăsirea unor imagini, estimarea similarității, identificarea caracteristicilor și analiza structurii temporale.

Multe instrumente au fost recent introduse în viața de zi cu zi pentru a capta sau pentru a edita conținutul digital, pentru a stoca și a distribui aceste înregistrări. În paralel cu această evoluție tehnologică, a apărut necesitatea unor sisteme de indexare care să permită filtrarea și regăsirea informațiilor relevante dintr-o masă de date.

¹ Profesorul Philippe Joly – Universitatea Toulouse III, IRIT, Institutul de Cercetări în Informatică din Toulouse, autor a acestui capitol, tradus de Mihaela Costin, colaborare dezvoltată în stagiul post-doctoral din cadrul proiectului POSDRU 56815, „Societatea bazată pe cunoaștere, cercetări, dezbateri, perspective”, la IRIT, Toulouse, Franța.

Inspirate din tehnologiile de indexare de text, cercetările actuale se concentrează pe indexarea automată a conținutului audiovizual. Ca suport al acestor cercetări, găsim o schemă generală ce corespunde posibilităților aplicative. Aceasta constă în crearea unui vector de trăsături descriptive mai mult sau mai puțin discriminative, extrase din fiecare unitate de conținut. Interogarea formulată de către un utilizator final, poate fi transformată într-un vector de trăsături. În versiunea simplă a acestei scheme, vectorii generați automat sunt în final confrunțați, comparați, cu termenul de interogare: o metrică, cum este o distanță sau o măsură a similarității este folosită pentru a efectua această comparație. Vectorii (și conținutul indirect indexat) sunt ordonați după distanța crescătoare.

În această schemă, există două etape în care mecanismele cognitive pot fi implicate. În primul rând, trăsăturile ce caracterizează conținutul pot fi inspirate în natura lor, sau în modul în care sunt extrase, din teoriile dezvoltate în domeniile percepției și cogniției, cunoașterii. Apoi, metrica folosită pentru a compara vectori poate fi mai mult sau mai puțin sofisticată, luând în considerare același tip de proprietăți. Multe lucrări au propus integrarea în procesul de estimare a distanței a unor parametri exogeni, cum ar fi aprecierea utilizatorului referitor la ceea ce înseamnă „similaritatea”, sau „sensibilitatea”, în percepția conținutului.

În paragrafele următoare, vom detalia acest subiect. Chiar dacă cele mai multe dintre aceste exemple provin din lucrările recent dezvoltate în cadrul grupului de cercetare SAMOVA al Institutului de Cercetări în Informatică din Toulouse, IRIT, Universitatea Paul Sabatier, în ultimul deceniu, trebuie însă să subliniem faptul că realizările din acest domeniu sunt în continuă expansiune. Ca urmare, ne vom concentra doar pe un număr redus de probleme și vom prezenta câteva dintre caracteristicile uzual folosite pentru indexarea imaginilor și a sunetului, din încercările făcute în a replica, pe cât posibil, comportamentele fiziologice și psihologice umane. Vom discuta apoi despre conceptul de similaritate și modul în care aceasta poate fi pragmatic integrat într-un motor de căutare. Pentru ilustrare, vom da detalii cu privire la unele aspecte din domeniul identificării audiovizuale a persoanelor, și la analiza structurii audiovizuale a unor înregistrări.

2. Clasificarea unor scene pentru regăsirea imaginilor

Având în vedere potențialul mare al domeniului exploatării conținutului digital cu caracter personal, s-au întreprins cercetări încă din timpul anilor '90, cu scopul dezvoltării sistemelor capabile să efectueze clasificarea automată a fotografiilor deși clasele erau destul de limitate și nu neapărat definite conform nevoilor utilizatorului final. În cele mai multe lucrări pe această temă, clasele sunt mai oarecum aceleași, cum ar fi, de exemplu, „peisaj urban”, „pădure”, „malul mării”, „munți”, „scene de interior” etc. Până în prezent, cercetătorii au încercat nu numai să extindă tipurile de clase, dar și să detecteze concepte care apar doar în anumite zone ale imaginilor, să găsească soluții în măsură să se confrunte cu dificultățile survenite din modificările conținutului imaginii, sau să utilizeze toate informațiile care pot fi extrase din surse externe (cum ar fi ontologiile). Pentru a compara rezultatele și a

stabili noi direcții de dezvoltare în acest domeniu, sunt organizate anual întruniri internaționale care permit cercetătorilor să confrunte metodele și rezultatele obținute lucrând pe aceleași seturi de date.

2.1. Caracteristicile (trăsăturile) globale

Când clasificare a fost limitată la detectarea câtorva clase, caracteristicile globale au constituit adeseori o soluție suficientă pentru clasificatori, pentru a obține rezultate de o bună calitate. Caracteristicile globale tipice caracterizează textura și culoarea într-o imagine.

Filtrele Gabor au reputat un succes deosebit, fiind printre primele introduse în descriptorii de tip MPEG-7, datorită proprietăților lor discriminatorii (amplificare, orientare) dar și posibilității de face legătura cu reprezentarea semantică. Pe lângă aceste caracteristici, multe lucrări au căutat o cale inspirată mai mult din biologie. Pe baza rezultatelor experimentale obținute în domeniul neuropsihologiei², au fost dezvoltate bănci de filtre pentru a modela felul în care se presupune că reacționează ochiul uman la conținutul diferit de texturi dintr-o imagine.

Alte lucrări încearcă să modeleze procesul uman de estimare al similarității între conținutul a două imaginii prin transformări dedicate ale caracteristicilor, sau prin agregări ale descriptorilor de textură^{3 4}.

O caracteristică foarte ușor discriminatorie și eficientă pentru indexarea de imagini este culoarea. Aceasta constituie un real suport în constituirea vectorilor de descriere prin trăsături, prezentând robustețe la numeroase transformări, cum ar fi modificările geometrice sau schimbările în intensitate luminoasă, prin selectarea ingenioasă a unui spațiu de culoare. Cum vectorii descriptivi trebuie să fie comparați între ei, un spațiu de culoare care respectă proprietățile perceptive (cum ar fi spațiul de culoare LAB⁵) este foarte adecvat. Cu toate acestea, în literatura de specialitate, această posibilitate nu a fost mult exploatată, poate pentru că păstrarea spațiului de culoare folosit pentru codarea inițială a imaginii, părea a fi o abordare cu costuri de calcul scăzute și suficient de eficientă⁶.

În general, un descriptor de culoare este de fapt o histograma de culoare calculată după o cuantificare a culorilor. Unul dintre neajunsurile majore ale unui descriptor global de culoare constă în dificultatea estimării similarităților. Există în principal două strategii de indexare. Pe de o parte, sunt generați vectorii tuturor imaginilor într-un prim pas de cuantizare, cu o aceeași hartă de culoare (adoptând o

² Jeanny Hérault, *De la rétine aux circuits neuromorphiques*, Chap, 3, in "Les systèmes de vision" J.M. Jolion ed., IC2 col., Hermes, Paris, 2001.

³ Hervé Le Borgne, Nathalie Guyader, Anne, Guérin-Dugué, Jeanny Hérault, *Classification of images: ICA Filters vs Human perception*. In Proc. Of ISSPA 2003.

⁴ Simon Thorpe, Arnaud Delorme, Rufin VanRullen and Williams Paquier (2000) *Reverse engineering of the visual system using networks of spiking neurons*. IEEE International Symposium on Circuits and Syst 4 405-408.

⁵ LAB – abreviere care provine din: L-luminozitate, A- axa a: red-green, B- axa b: blue-yellow (nota editorului).

⁶ Rezultate deosebite în cercetări foarte recente pe aceeași direcție au fost obținute chiar în cadrul colectivelor de cercetare ale Academiei Române – nota editorului.

anume „paletă” de culori). Pe de altă parte, cuantificarea, cuantizarea, se realizează referitor la culoarea care apare efectiv în fiecare imagine (realizând o ordonare, o aglutinare a culorilor). Atunci când în urma primei abordări, toate imaginile sunt indexate cu un același tabel de cuantizare, vectorii pot fi comparați rapid, cu metode relativ simple (cum ar fi distanța Manhattan).

Dar, în cazul unor devieri cromatice, chiar și mici, vectorii de culoare pot fi foarte diferiți și imaginile vizual foarte similare pot fi clasificate foarte diferit, prin acest mecanism. Într-o a doua abordare, sunt folosite distanțe mai sofisticate care să ia în considerare atât vectorii, cât și tabelele de cuantificare (calculând distanța pătratică sau EMD – Earth Mover Distance – care este o măsură a distanței dintre două distribuții de probabilitate pe o regiune D). Se estimează rezultate astfel obținerea unor rezultate mai bune, dar costul de calcul reduce puternic cantitatea de comparații ce pot fi realizate într-un timp interactiv rezonabil.

Procesul de cuantizare a culorii este o problemă-cheie în cele mai multe aplicații de indexare de imagine. În multe cazuri, un descriptor de tip „culoare dominantă” este folosit drept trăsătură distinctivă, dar modul în care culorile dominante sunt extrase este în general discutabil. Culorile dominante pot fi definite ca fiind corespunzătoare primelor N vârfuri din maximele din histograma de culoare (care este mai mult sau mai puțin și propunerea făcută în [MPEG-7]).

Deoarece într-o imagine naturală am putea avea doar pixeli de culori diferite, este necesară mai întâi o cuantificare, pentru a putea extrage culorile dominante. Dacă luăm în considerare răspunsurile date de utilizatorii finali atunci când li s-a cerut să indice culoarea dominantă pentru o imagine dată, am putea observa diferențe importante față de rezultatele date de un proces automat. De obicei, o astfel de diferență apare pentru imagini care prezintă un singur obiect pe un fundal uniform. Când computerul identifică culoarea de fundal ca fiind cea dominantă, utilizatorii se uită în general doar la obiect pentru a determina ce culoare dominantă are. Evident, în cele mai multe cazuri, identificarea unei similarități puternice între două imagini care prezintă același obiect pe două fundaluri diferite este un rezultat considerat satisfăcător.

Acest neajuns nu este specific doar descriptorilor de culoare, ci, în general, „descriptorilor globali”, care caracterizează întreaga imagine și care nu se concentrează pe arii specifice din imagine. „Descriptorii globali” pot fi ușor transformați în descriptori locali prin divizarea unei imagini în regiuni mai mici, printr-un algoritm de extragere care aplică aceeași funcție pentru fiecare dintre aceste zone. Unii descriptori au fost, mai recent, special concepuți pentru a surprinde caracteristici discriminatorii locale, ale unei imagini.

2.1. Caracteristicile (trăsăturile) locale

Generarea automată a hărților de puncte care atrag în mod preponderent atenția, saliente, (SM „saliency map”) presupune evidențierea pozițiilor dintr-o imagine sau dintr-un videoclip în care ochiul ar trebui să insiste (să zăbovească) mai mult timp atunci când se parcurge conținutul imaginii. SM harta saliențelor sau „harta atenției” cum ar mai putea fi numită, poate fi produsă de un dispozitiv de

urmărire a privirii. Dar în domeniul indexării automate a conținutului, obiectivul este de a genera astfel de hărți utilizând modele predictive, în general bazate pe textura locală și conținutul de mișcare. Ideea principală este că informațiile discriminatorii sau semantice nu sunt omogen distribuite în imagine. Prin urmare, există zone mai relevante, care trebuie descrise prin mai multe detalii, în procesul de indexare.

Generarea automată a acestor hărți SM ar putea fi inspirată din domeniul neuropsihologiei dar asta se întâmplă rareori în cercetarea clasică care nu ia în considerație elemente noi, din domeniul științelor cognitive, deoarece rezultatele sunt evaluate în funcție de precizie pentru documentul returnat de un motor de căutare, și nu în funcție de zonele saliente, de harta de zone relevante. Dintre lucrările recente, putem menționa aici „zoomable interactive video” sistem dezvoltat de Vincent Charvillat⁷, în ultima decadă.

Scopul urmărit este de a mixa o primă hartă de zone saliente care indică prioritățile calculate pentru zonele cu un nivel ridicat al gradientului de densitate, sau care identifică obiecte în mișcare, cu o altă hartă de zone saliente construită prin reținerea locațiilor unor clicuri efectuate de utilizatori când doresc să mărească anumite părți specifice dintr-un video.

Datorită acestor cunoștințe agregate, schema de codare a înregistrării video poate fi adaptată pentru a păstra anumite detalii la o rezoluție înaltă în regiunile de interes (ROI „regions of interest”), cu scopul de a oferi posibilitatea unui zoom eficient, păstrând în același timp o rată de compresie rezonabilă.

S-a propus marcarea unor puncte de interes pentru realizarea unui model al zonelor unde privirea persistă un timp foarte scurt înainte de a trece mai departe, atunci când este explorată vizual imaginea. După cum s-a menționat deja, în domeniul indexării multimedia, cercetătorii au acordat atenție mai ales robusteții și invarianței algoritmului de extracție dezvoltat, sau proprietății discriminative a descriptorului atașat fiecărui punct. Am putea spune însă că punctele și descriptorii SIFT și SURF sunt departe de considerațiile cognitive. Descriptorii uzuali atașați punctelor de salience utilizează distribuția gradientului în vecinătatea fiecărui punct (și anume „histograme de gradient” sau „HoG”). Având în vedere puterea discriminativă remarcabilă a acestor descriptori, s-a propus utilizarea lor ca elemente pentru un nou tip de schemă de codare a imaginii. Ideea este de a colecta regiunile din imagini care generează puncte similare de interes (cu descriptori similari), pentru a putea duce aceste regiuni pe o imagine inițial albă (vidă), și utilizând metodele de inpainting și de interpolare să se umple zonele libere rămase goale, să se genereze o nouă imagine. Deoarece imaginile obținute cu acest proces invers sunt de o calitate suficient de bună, putem vedea descriptorii inițial concepuți ca fiind semnături discriminatorii eficiente, ce pot servi la reprezentarea prin compresie a unui conținut vizual. Mai mult decât atât, putem trage concluzia că ariile înconjurând aceste puncte de interes (SIFT sau SURF) sunt corespunzătoare părților semantice semnificative din conținutul vizual al înregistrării.

⁷ Vincent Charvillat, Wei Tsang Ooi, Romulus Grigoraș, Géraldine Morin, *Crowdsourced Automatic Zoom and Scroll for Video Retargeting* în ACM Multimedia, Florence, 25/11/2010-29/11/2010, ACM, (suport electronic), 2010.

2.2 Trăsături audio

Atunci când o coloană sonoră a fost înregistrată împreună cu conținutul vizual care urmează a fi indexat, s-a observat o complementaritate între indicii care pot fi extrași prin fiecare dintre aceste modalități. De exemplu, complexitatea problemei identificării unei persoane într-un context audiovizual poate fi mult redusă dacă se folosesc în același timp atât indicii audio cât și cei video.

Utilizarea indicilor audio apare astăzi ca o sursă alternativă de informare dacă dorim să clasificăm automat locația unei înregistrări. Putem înțelege evident că un sunet de valuri poate ajuta la identificarea unei scene de plajă, sau un sunet de claxon sau de motor poate fi de util pentru a decide dacă un videoclip ilustrează imaginea unei străzi. Trăsăturile audio pot fi mai relevante în unele cazuri decât cele video pentru identificarea anumitor evenimente, cum ar fi „spălarea unor vase”, înregistrată de dispozitivele de monitorizare folosite pentru asistarea persoanelor în vârstă la domiciliu⁸.

Pentru a procesa coloana sonoră a documentului audiovizual, sunt uzual extrase câteva caracteristici, pentru a caracteriza unele elemente specifice conținutului, cum sunt vorbirea, muzica, aplauzele, râsetele, sunetul de clopoței, ... etc.

Pentru a procesa coloana sonoră a documentului audiovizual, câteva caracteristici sunt uzual extrase, pentru a caracteriza unele elemente specifice conținutului, cum sunt vorbirea, muzica, aplauzele, râsetele, sunetul de clopoței, ... etc.

Aceste caracteristici sunt:

- coeficienții spectrali (sau cel puțin magnitudinea lor);
- coeficienții cepstrali⁹;
- „energia de modulare la 4 Hz”, care caracterizează viteza silabică de exprimare înregistrată la 4 Hz;
- modularea entropiei care caracterizează nivelul de organizare al conținutului muzical etc.

Ca și în cazul descriptorilor vizuali, MPEG-7 oferă un mare număr de descriptori audio utili pentru caracterizarea conținutului audio¹⁰ (cum ar fi descriptorul anvelopă spectrală, sau împrăștierea spectrală).

Pentru cele mai multe dintre încercările noastre am utilizat frecvent rezultatele directe sau indirecte ale unei metode de segmentare audio numite „forward-backward segmentation”¹¹, segmentare bazată pe propagarea „forward-backward”

⁸ Patrice Guyot, Julien Piquier, Régine André-Obrecht, *Water Flow Detection From a Wearable Device With a New Feature, The Spectral Cover* in International Workshop on Content-Based Multimedia Indexing (CBMI 2012), Annecy, France, 27-29/06/2012, IEEE, pp. 139-142, June 2012.

⁹ Denumirea de *cepstru* provine din citirea inversă a literelor primei silabe din cuvântul *spectru* (nota editorului); deci este o transformare inversă a spectrului, Transformata Fourier a log din magnitudinea spectrului Fourier.

¹⁰ B. S. Manjunath, Philippe Salembier, Thomas Sikora, *Introduction to Mpeg-7: Multimedia Content Description Interface*. Editor : John Wiley & Sons Ltd, 2002.

¹¹ Régine André-Obrecht. *A new statistical approach for automatic speech segmentation*, in Transactions on Audio, Speech, and Signal Processing, IEEE, vol. 36, nr. 1, pp. 29-40, 1988.

(înapoi), cu scopul de a diviza semnalul audio în unități cvasi-staționare caracterizate prin modele gaussiene cu autoregresie.

Astfel, putem astfel observa numărul de segmente generate într-o fereastră alunecătoare (care este în general mai mic pentru un conținut muzical decât pentru un conținut vorbit, rostit), și durata segmentelor (care sunt de obicei mai lungi în cazul muzicii decât al vorbirii) etc.

Unele caracteristici pot fi utilizate pentru a încerca să ne apropiem de modul de funcționare a sistemului auditiv uman. Se poate folosi, de exemplu, un banc de filtre Gammatone pentru a genera un vector descriptiv. Filtrele Gammatone sunt definite ca produsul a doi factori: o funcție cosinus și o distribuție gamma:

$$g(t) = a \cdot t^{n-1} \cdot e^{-2\pi b t} \cdot \cos(2\pi f t + \phi) \quad (1)$$

Aceste caracteristici au fost comparate cu unele caracteristici clasice (coeficienții cepstrali și descriptorii MPEG-7) pentru evaluarea performanțelor lor în geolocalizarea sunetelor înregistrate¹². În aceste lucrări s-a folosit o bază de date de sunete diferite pentru antrenare clasificatorilor. Evaluarea rezultatului se realizează prin calculul distanței medii până la locația efectivă unde sunetul testat a fost înregistrat. În rezultatele studiilor actuale, coeficienții cepstrali și filtrele Gammatone generează o descriere cu aproape aceeași abilitate de a caracteriza conținutul, în timp ce descriptorii MPEG-7 par a fi un pic mai eficienți pentru o astfel de activitate. Studiul nu contestă interesul pentru filtrele Gammatone dar sunt necesare noi cercetări pentru a determina schemele de decizie în care aceste filtre pot fi folosite.

Unul din procesele principale în analiza automată a unei coloane sonore (pentru filme și conținut TV) constă în a determina prezența a patru clase de sunete: vorbire, muzică, sunete și tăcere; ulterior, fiecare clasă poate fi împărțită în subclase. În cazul discursului, este necesară segmentarea conținutului în funcție de vorbitori (Guyot 2012)¹³, și pe baza acestor rezultate să se determine rolul jucat de fiecare vorbitor (Manjunath 2002)¹⁴. În cazul muzicii, se dorește separarea conținutului vocal de cel instrumental, cât și clasificarea după criteriul monofonic/polifonic (o singură voce la un moment dat sau mai multe voci în simultaneitate). Una dintre caracteristicile de bază pentru un astfel de proces constă în extragerea „pitch-ului” sau a frecvenței fundamentale.

Am putea apela la definițiile din domeniul fizicii pentru a determina metodele de extragere ale acestor caracteristici, dar din punctul de vedere al percepției, acest proces s-ar putea confrunța cu unele neajunsuri. Acesta este motivul pentru care pitch-ul este extras prin algoritmi mai mult sau mai puțin sofisticati, cum ar fi algoritmul Yin¹⁵ bazat pe funcția „diferență normalizată medie cumulativă”.

Sunetele sunt deseori utilizate în industria cinematografică pentru a transporta (a furniza) informații diegetice. Scopul constă în adăugarea la conținutul

¹² Xavier Sevillano, Xavier Valero, Francesc Alias, *Audio and video cues for geo-tagging online videos in the absence of metadata*, in proc. of CBMI 2012, pp. 217-222.

¹³ Patrice Guyot, Julien Pinquier, Régine André-Obrecht, *Idem*.

¹⁴ B. S. Manjunath, Philippe Salembier, Thomas Sikora, *Idem*.

¹⁵ Alain de Cheveigné, Hideki Kawahara, *YIN, a fundamental frequency estimator for speech and music*, in Journal of Acoustical Society of America, 111 (4), pp. 1917-1930, april 2002.

audiovizual a unor elemente referitoare la mediu, care ar putea da unele informații despre locația sau persoanele din jur, sau despre unele obiecte virtuale (primăvară, elicopter etc). Cu alte cuvinte, aceste elemente sonore constau în sunete care corespund unor surse care nu pot fi văzute pe ecran. Unele experimente arată că publicul relatează că a și văzut ceea ce de fapt a fost perceput doar în urma înregistrărilor de pe coloana sonoră¹⁶. Acest principiu este o limitare majoră în corelarea sursei de sunet cu un concept vizual real și duce la unele probleme cum ar fi localizarea automată a sursei sunetului (in / out / off). Pentru această problemă, primul pas constă în localizarea fețelor (a chipurilor) și în determinarea așa-numitei „rate de activitate a buzelor”.

Această caracteristică poate fi apoi folosită pentru a determina dacă discursul detectat pe coloana sonoră poate realmente corespunde unei persoane vizibile pe ecran¹⁷.

3. Estimarea similarității

Odată ce un conținut este descris prin caracteristicile extrase, următoarea problemă de rezolvat este să se estimeze similaritatea conținutului. Această informație are sens doar dacă poate fi corelată cu o scală umană de percepție a similarității. N-am putea întreba care este modalitatea prin care creierul uman reușește să estimeze similaritatea între două melodii? Trebuie să luăm în considerație trăsăturile locale sau globale, să le comparăm, pentru a determina valoarea acestei similarități?¹⁸

Procesul în sine trebuie să se încadreze în anumite constrângeri, ca de exemplu un calcul rapid al distanțelor în cazul unui motor de căutare. Acest obiectiv poate fi atins realizând o cuantificare semnificativă și metode de calcul rapid. În acest caz, scorurile de evaluare, rata de precizie și de recall (recunoaștere), sunt mai mici decât cele care se pot atinge cu algoritmi de cuantificare adaptivi de precizie, luând în considerare unele proprietăți perceptuale, care însă necesită instrumente de estimare a distanței, costisitoare (cum ar fi „Fuzzy-C means” pentru cuantificare, distanța pătratică sau Earth Mover Distance pentru estimarea similarității).

Când utilizăm matrice de similaritate^{19, 20}, pentru a compara documentele temporale (sub aspectul conținutului audio sau video), formulăm o ipoteză implicită considerând că similaritatea este un concept care ar trebui să ia în considerare atât conținutul cât și structura, în funcție de timp. Câțiva algoritmi au fost propuși pentru a efectua acest tip de comparație cu diverse noi aplicații cum ar fi estimarea

¹⁶ Michel Chion, *L'Audio-Vision, son et image au cinéma*, Ed. Nathan Université, 1991.

¹⁷ Elie El Khoury, Christine Senac, Julien Pinquier, *Improved Speaker Diarization System for Meetings*. In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009), Taipei, Taiwan, 19/04/09-24/04/09, IEEE, pp. 4241-4244, 2009.

¹⁸ Geoffroy Peeters, Xavier Rodet, *Signal-based Music Structure Discovery for Audio Summary Generation*, ICMC03 Singapore, Singapore 2003 29 Sep - 4 Oct.

¹⁹ Matthew Cooper, and Jonathan Foote, *Summarizing Popular Music via Structural Similarity Analysis*, in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2003).

²⁰ Bertrand Delezoide, *Hierarchical film segmentation using audio and visual similarity*, in Proceedings of the IEEE International Conference on Multimedia & Expo (ICME'05), 2005.

„similarității stilului” pentru regăsirea unui conținut în imagini, sau segmentarea temporală pentru a împărți o înregistrare lungă în programe sau în secvențe.

În lucrările precedente s-a considerat că similaritatea stilistică se bazează în principal pe apariția elementelor comune în documentele comparate, referitor la modul în care au fost generate, produse. Aceste elemente comune, din punct de vedere perceptiv uman, denumite invariante de producție „production invariants”, sunt uneori asociate cu o combinație tipică de caractere audio-vizuale. O astfel de combinație poate evidenția faptul că o culoare dominantă (care corespunde la un anumit set dat, la o iluminare dată), evoluează în mare în același mod pe parcursul a două documente diferite, sau, la un alt nivel, faptul că o reclamă se repetă la momente diferite într-un program TV. Măsurarea gradului de similaritate stilistică dintre două înregistrări se bazează pe abilitatea de a cuantifica evoluția similară a unui subset de elemente în anumite momente.

Să considerăm de exemplu că vrem să comparăm două secvențe temporale S1 și S2 (Fig. 1). Extragem pentru fiecare secvență câțiva vectori de trăsături la intervale regulate de timp: trei vectori pentru prima secvență, cinci pentru a doua. Fiecare vector este format din 2 valori $[v1, v2]$ corespunzătoare la unele măsuri privind conținutul (luminanța medie și energia sonoră, de exemplu). În matrice, vom identifica doar trăsăturile vectorului care iau aceeași valoare.

S2	$\begin{bmatrix} 1 \\ 4 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 4 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 3 \end{bmatrix}$	$\begin{bmatrix} 3 \\ 4 \end{bmatrix}$
S1					
$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$	f1		f1 & f2		
$\begin{bmatrix} 2 \\ 3 \end{bmatrix}$		f1		f1 & f2	
$\begin{bmatrix} 3 \\ 4 \end{bmatrix}$	f2	f2			f1 & f2

Fig. 1. Compararea a două secvențe

Când două secvențe vectoriale partajează (au în comun) o sub-componentă similară, ea poate fi evidențiată printr-un segment pe diagonală unde câteva trăsături (de sperat toate) au aceleași valori de-a lungul timpului. Dacă obiectivul este de a identifica sub-secvențe audio sau video similare din lungi înregistrări, o soluție ar putea fi extragerea unui set de caracteristici, la intervale de timp regulate, pentru a calcula această matrice de similaritate, și căutarea segmentelor diagonale pentru localizarea fragmentelor similare.

Dar trăsăturile ar putea fi analizate și ca serii de timp. Pentru a măsura similaritatea dintre două secvențe, una din cele mai populare metode folosite este LCSS (Longest Common SubSequence care ar putea fi tradusă prin cel mai lung subșir comun, cea mai lungă sub-secvență comună) și DTW (Dynamic Time Warping – care ar putea fi aproximativ tradusă drept similaritatea înfășurătoare în

timp calculate dinamic) și variante ale acestora. Câteva lucrări au propus utilizarea unei metrici de similaritate ca instrument de comparație. Principala problemă în această abordare rezidă în costul computațional. Pentru a reduce acest cost am putea genera o matrice de similaritate aproximativă²¹, sau am putea aproxima rezultatul unui proces de tip DTW²².

Pentru a produce o măsură simplă a similarității între două documente audiovizuale, se poate folosi următoarea metrică:

$$\text{Max}_{k=1}^M \left(\sum_{i=1}^N w_{ki} \cdot \left(\frac{\sum \text{coef on the } i^{\text{th}} \text{ diagonal}}{\text{Nb of coef on the } i^{\text{th}} \text{ diagonal}} \right) \right) \quad (1)$$

unde i și k sunt indici utilizați pentru a parcurge respectiv toate diagonalele N ale matricei de similaritate și toate dimensiunile vectorului M și termenii w_{ki} corespund funcției de ponderare care poate fi aleasă în funcție de aplicație pentru a da o importanță mai mare sau mai mică sincronizării dintre cele două șiruri, sau unor trăsături care prezintă unele proprietăți interesante pentru segmentarea secvențelor în zone specifice, de exemplu.

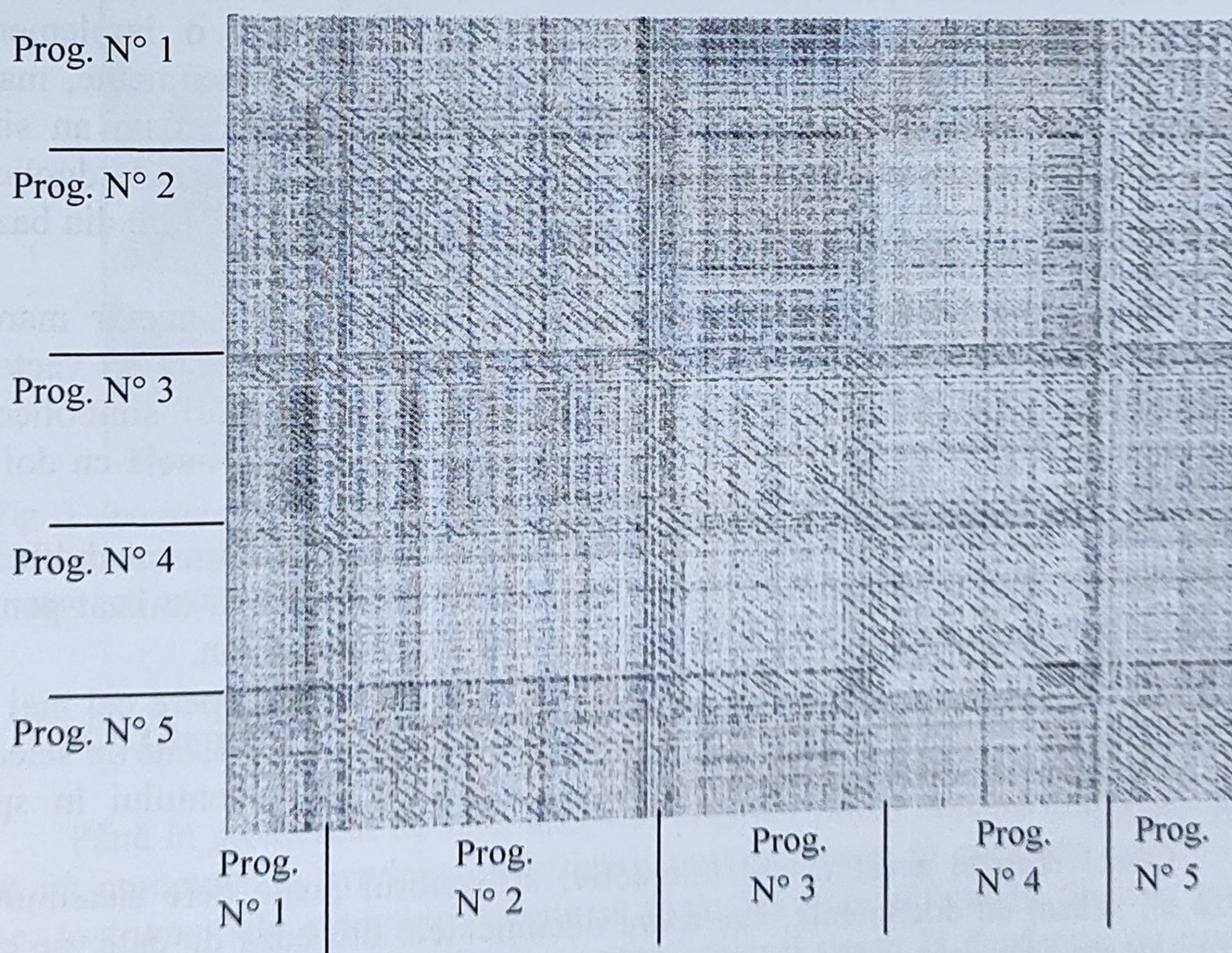


Fig. 2. Structura de blocuri pusă în evidență de către matricea de similaritate 2x4 atunci când se compară aproximativ două zile consecutive de televiziune. Fiecare bloc corespunde unui program TV.

²¹ Siba Haidar, Philippe Joly, Bilal Chebaro, *Mining for Video Production Invariants to Measure Style Similarity*, in International Journal of Intelligent Systems, Wiley, vol. 21, nr. 7, pp. 747-763, July 2006.

²² Romain Tavenard, Laurent Amsaleg, Guillaume Gravier, *Model-based similarity estimation of multidimensional temporal sequences*, in Annales des Télécommunications 64, 5 (2009) 381-390.

Această metrică poate fi utilizată pentru a identifica un fel de „similaritate stilistică” între două înregistrări video. Urmând aceeași observație, o matrice de similaritate poate fi utilizată și pentru a evidenția meta-structurile în fluxuri video importante, de exemplu, atunci când ne uităm la matricele obținute pe înregistrările TV făcute timp de două zile pe aceleași canale, observăm unele blocuri largi, fiecare aparținând unui singur program TV.

În cele mai multe dintre lucrările dezvoltate pe tema estimării similarității conținuturilor audiovizuale ale înregistrărilor, doar caracteristicile vizuale sau audio extrase în mod automat sunt utilizate pentru a calcula metricile dintre vectorii descriptivi.

Această abordare se bazează pe un punct de vedere arbitrar, unde utilizatorul final nu este luat în considerare. În funcție de context, de sarcina de efectuat, sau de alte aspecte, putem fi mai mult sau mai puțin sensibili la importanța unor aspecte date. De exemplu, în cadrul laboratoarelor japoneze au fost făcute unele propuneri, cu scopul de a ține seama de sensibilitatea utilizatorului la background, în estimarea similarității²³, (fenomen denumit „kansei”).

Pentru a integra percepția utilizatorului final despre ceea ce ar însemna similaritatea dintre două conținuturi audiovizuale, am propus o implementare interactivă²⁴ în care organizarea vizuală a unui mic sub-set de documente, manual definită de către un utilizator, este considerată ca o intrare pentru un an sistem supervizat care încearcă să găsească metrica și parametrii care să permită duplicarea aceleiași organizări spațiale localizând astfel automat fiecare înregistrare din baza de date.

Primul pas din acest process constă în extragerea unui număr mare de trăsături de nivel scăzut din înregistrarea audio-video, care formează un vector de descriere pentru fiecare document. Odată ce primele reprezentări simbolice ale documentelor selectate sunt plasate pe ecran, sistemul intră într-o buclă cu doi pași până când apare o comportare staționară a rezultatului:

- **Pasul 1.** Selectarea unui sub-set de trăsături printr-un algoritm cu dublu sens (Forward Backward Feature Selection algorithm) care poate fi utilizat pentru a calcula o distanță normalizată între documentele plasate pe ecran.
- **Pasul 2.** Algoritm cu regresie (ϵ -SVR) care încearcă să descopere cei mai buni parametri cu scopul de a potrivi distanța normalizată dintre trăsăturile selectate, cu distanța euclidiană efectivă dintre reprezentările documentului în spațiul ecranului.

Odată realizat acest calcul interactiv, utilizatorul poate cere sistemului să afișeze un subset de documente, sau *toate* documentele din baza de date, pe ecran. Pentru a depăși constrângerile care nu pot fi satisfăcute prin proiecția 2D a spațiului vectorilor descriptivi pe ecran, considerăm entitățile vizuale ca fiind niște particule

²³ Sunkyoung Baek, Miyoung Cho, Myungwon Hwang, Pankoo Kim, *KANSEI-Based Image Retrieval Associated with Color*, in Fuzzy Logics and Applications, Lecture Notes in Computer Science, Volume 3849, pp. 236-333, 2006.

²⁴ Jeremy Philippeau, Julien Pinquier, Philippe Joly, Jean Carrive, *Dynamic organization of audiovisual database using a user-defined similarity measure based on low-level features*, in IEEE International Conference on Image Processing (ICIP 2008), San Diego, California, U.S.A., 12/10/08-15/10/08, IEEE, pp. 33-36, 2008.

fizice și implementăm un model fizic dinamic, de tip sistem oscilator armonic „resort-masă”, denumit „mass spring”.

Am implementat acest model cu un algoritm Runge-Kutta de ordinal patru, o schemă de integrare explicit temporală cunoscută a fi foarte exactă și cu un bun comportament pentru o gamă extinsă de probleme. Setul global de documente este considerat un graf complet ale cărui noduri sunt particule ponderate și muchiile sunt arcurile. Rezultatul este o reprezentare dinamică unde reprezentările simbolice ale documentelor se mișcă încet către poziția lor finală, „împingând”, dacă este necesar, alte documente mai puțin similare, care stau în calea lor.

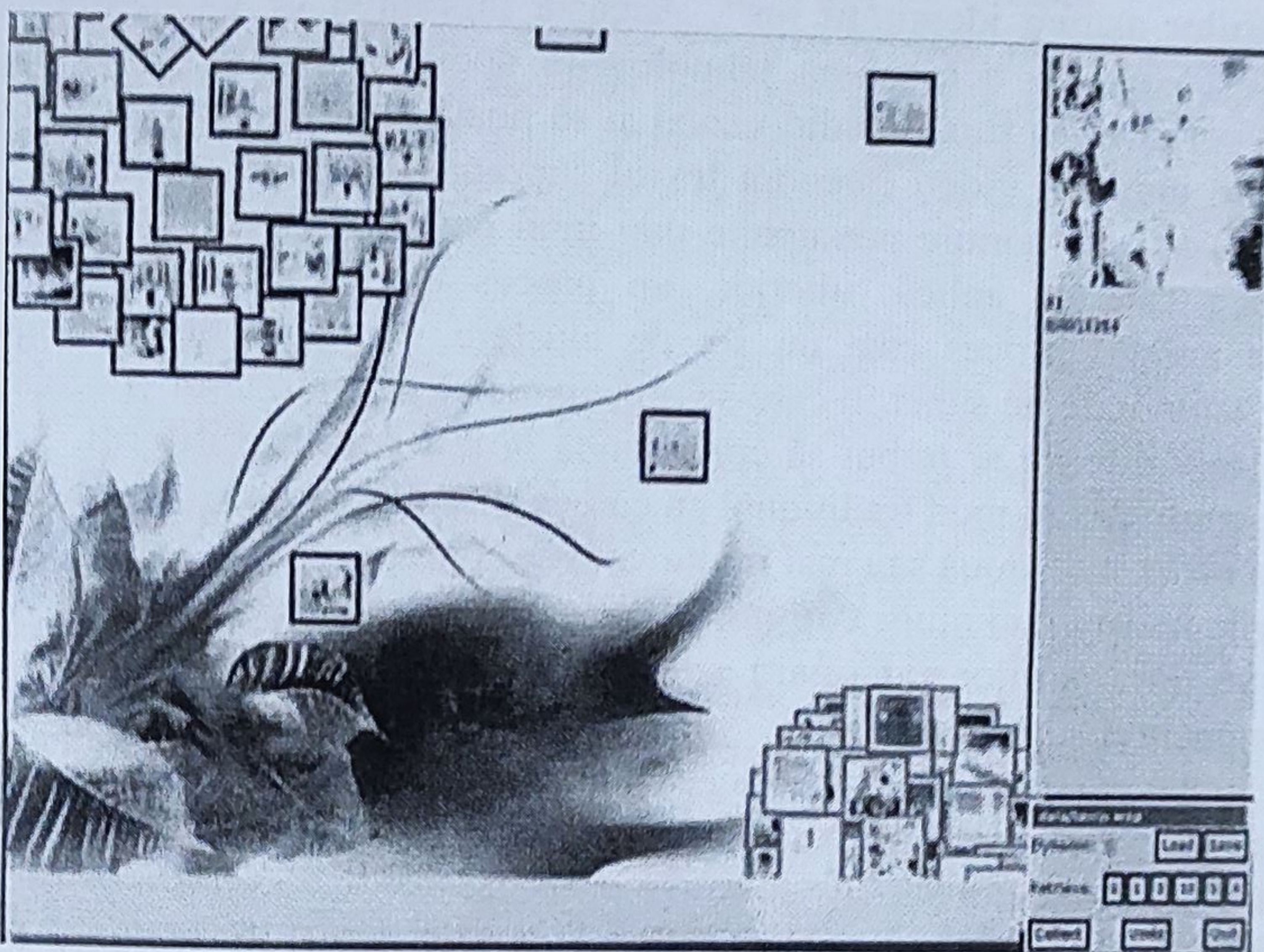


Fig. 3. Interfața aplicației pentru organizarea asistată a bazelor de date audio-vizuale. Pătratele din partea stângă sunt reprezentări simbolice ale fișierelor video. Organizare spațială automat extrapolată din elementele de interacțiune cu utilizatorul (copyright INA).

4. Identificarea persoanei

Până în momentul de față, sistemul cognitiv uman rămâne mult mai eficient decât un computer în privința identificării unei persoane în contexte audio-vizuale tipice. Instrumentele automate trebuie să facă față unor dificultăți diverse, precum rezoluția slabă sau orientarea feței, pe care creierul uman le depășește cu ușurință. Când se ia în considerare doar aspectul vizual al feței, una dintre cele mai eficiente abordări este „modelarea aspectului”. Ideea acestei abordări este să se cartografieze nodurile pe o rețea predefinită „mesh” (pe suprafața feței) și să se determine transformările minime pentru a identifica o persoană cunoscută. Lucrările recente care tratează procesul identificării persoanei abordează analiza conținutului audiovizual integrând trăsăturile acustice și vizuale. Scopul aplicațiilor este de a găsi, plecând de la modele vocale și fizionomice, secvențe în care apare o anumită persoană, fiecare figură fiind unic asociată unei voci. Totuși, în anumite aplicații,

asocierea audio-video nu este disponibilă. De pildă, pentru a procesa diverse difuzări de televiziune, nu dispunem de un model anterior și de aceea modelele trebuie calculate ad-hoc, în momentul apariției persoanei. Identificarea persoanei în înregistrări audio-vizuale reprezintă de mult timp o provocare, în primul rând pentru că acoperă niște aspecte ce țin de securitate. Recent, a apărut o tehnologie interesantă de a valoriza arhivele audio-vizuale prin intermediul unor funcționalități care permit să se restituie prima apariție a unei persoane sau să se realizeze statistici despre intervențiile acesteia (intervențiile unui politician, de exemplu). Obiectivul actual este identificarea personelor care apar în emisiunile televizate fără existența apriorică a informațiilor asupra identității lor²⁵. Aceasta înseamnă că cercetătorii trebuie să găsească mijloace de a combina tehnologiile audio și video pentru a detecta și localiza o persoană în fluxul audio-vizual și să poată extrage din coloana sonoră sau din anumite indicații video (precum textul suprapus) numele persoanei respective. Procesarea coloanei sonore presupune mai întâi detectarea segmentelor de conținut rostit, peste care se aplică, ulterior, un proces de „diarizare a vorbitorului”²⁶. Rezultatul acestui proces este un set de clustere. Fiecare cluster grupează niște segmente temporale de sunet în care se presupune că se aude unul și același vorbitor. Două clustere diferite ar trebui să corespundă la doi vorbitori diferiți. Acest gen de proces se confruntă în mod tradițional cu câteva dificultăți notorii, și anume:

- segmentele unde două sau mai multe voci se suprapun;
- cazurile în care mai mulți vorbitori au aproape aceeași voce;
- absența informațiilor anterioare asupra numărului de vorbitori diferiți care apar în fragmentul supus procesării.

Când se realizează manual aceeași analiză, dar fără înregistrarea imaginii, un subiect uman poate fi mai eficient în privința primului aspect, dar va resimți și el dificultatea celorlalte două.

Principalele strategii utilizate vizavi de algoritmul de formare a clusterelor constau în folosirea ca regulă decizională a relației de probabilitate generalizată (GLR – „Generalized Likelihood Ratio”) sau a variației criteriului inferenței bayesiene (Δ BIC „variation of the Bayesian Inference Criterion”) aplicate unui set de trăsături audio tradiționale (variația energiei la 4 Hz „4 Hz Energy Variation”, viteza de trecere prin zero „Zero Crossing Rate”, coeficienții mel-cepstrali MFCC – „Mel Frequency Cepstral Coefficients” etc)^{27,28}.

Δ BIC se definește după cum urmează:

²⁵ Juliette Kahn, Olivier Galibert, Ludovic Quintard, Matthieu Carré, Aude Giraudel and Philippe Joly. *A presentation of the REPERE Challenge*, in Proc. of CBMI 2012, pp. 102-107, Annecy June 27-29, 2012.

²⁶ Elie El Khoury, Christine Senac, Julien Piquier, *Improved Speaker Diarization System for Meetings*, in IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009, Taipei, Taiwan, 19+24/04/09, IEEE, pp. 4241-4244, 2009.

²⁷ Daoudi Khalid, Jérôme Louradour (2009), *A Comparison between sequence kernels for SVM speaker verification*, 4241-4244. In Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP).

²⁸ Jean-Luc Rouas, Jérôme Farinas, François Pellegrino, Régine André-Obrecht, *Rhythmic unit extraction and modelling for automatic language identification*, in Speech Communication, Elsevier, Vol. 47 N. 4, pp. 436-456, 2005.

$$\Delta BIC = (n_1 + n_2) \log |\Sigma| - n_1 \log |\Sigma_1| - n_2 \log |\Sigma_2| - \frac{\lambda}{2} \left(d + \frac{d(d+1)}{2} \right) \log (n_1 + n_2) \quad (2)$$

unde n_1 și n_2 sunt mărimile a două clustere testate care urmează să fie fuzionate; Σ_1 , Σ_2 și Σ sunt matricele de covarianță a celor două clustere luate individual sau combinate, fuzionate, d este dimensiunea vectorilor de trăsături, iar λ este un parametru optimizat.

Inspirați de lucrările făcute pe coloana sonoră (banda audio), ne putem referi acum la unele contribuții în „diarizarea” fețelor, a figurilor umane.

Obiectivul este același, iar procesul urmează aproape identic același traseu.

Fețele ce apar pe banda video sunt detectate, urmărite și marcate. Fiecare față urmărită conduce la identificarea unei „urme a feței” sau „marcaj fizionomic” („face track”)²⁹ și fiecare față nou marcată conduce la identificarea unei persoane. Sunt apoi extrași descriptorii din marcajele fizionomice în scopul de a operaționaliza un algoritm de formare a clusterelor al cărui obiectiv este să grupeze în cadrul aceleiași cluster toate marcajele fizionomice care corespund unei singure persoane.

În funcție de contextual aplicativ, construirea unui model fizionomic al persoanei poate necesita un proces de preselecție a imaginilor disponibile spre a le păstra doar pe cele mai relevante, mai ales atunci când o figură cheie trebuie să fie selectată pe o înregistrare, prin urmărire fizionomică (face track). Aceeași problemă apare, în cazul imaginilor returnate de un motor de căutare atunci când se caută o persoană celebră. Este posibil ca nu toate aceste imagini să corespundă cerințelor minime de antrenare a unui program de recunoaștere. S-ar putea să ni se returneze și desene sau imagini cu efect artistic care fac persoana de nerecunoscut.

În astfel de situații considerăm că este preferabil să se păstreze doar imaginile în care:

- fața este cât mai mare cu putință: cu cât rezoluția este mai bună, cu atât mai precis va fi modelul;
- fața este cât mai vizibilă, cu cât rezoluția este mai bună, cu atât mai precis va fi modelul;
- fața nu este parțial acoperită de părul lung, de vreun obiect sau de vreun efect de iluminare;
- fața este poziționată frontal și orientată vertical – această condiție nu este obligatorie pentru toate instrumentele de recunoaștere, totuși unora dintre ele trebuie să li se prezinte doar fețe în posturi normalizate.

Pentru a selecta automat imaginile unde acele proprietăți sunt respectate, propunem să utilizăm patru factori normalizați calculați pe regiuni din figură, și anume: funcția care dă o parte din suprafața de piele – RSP („ratio of skin part”), mărimea feței – FS („face size”), raportul feței – FR („face ratio”) și orientarea feței – FO („face orientation”)³⁰.

²⁹ Mark Everingham, Andrew Zisserman. *Automatic Person Identification in Video*, in Image and Video Retrieval, vol. 3115. LNCS 2004.

³⁰ Ikram Ben Kouas, Philippe Joly, *Automatic Keyframe Selection for Known People Identification in Images*, in Proc. of ECMS 2011. IEEE Explore, 2011.

Definim RSP („ratio of skin part”) ca fiind numărul de pixeli care corespunde pielii, raportat la numărul total de pixeli ai unei zone rectangulare selectate pentru delimitarea feței. Clasificarea pixelilor în clasele skin și not-skin poate fi realizată cu instrumente euristice simple aplicate asupra trăsăturilor de cromatică.

FS („face size”) este o mărime dată de formula $\frac{1}{w \cdot h}$ unde w și h sunt respectiv lățimea și înălțimea zonei rectangulare de delimitare.

Valoarea ideală a raportului dintre lățimea zonei dreptunghiulare de delimitare și înălțime, este de $\frac{3}{4}$. Valoarea FR rezultă utilizând raportul dintre dimensiunea ideală și cea reală.

$$FR = 1 - \frac{\min\left(\frac{w}{h}, \frac{3}{4}\right)}{\max\left(\frac{w}{h}, \frac{3}{4}\right)} \quad (3)$$

FO este o mărime a cărei valoare caracterizează aspectul simetric al feței în vedere frontală și orientarea sa verticală, calculându-se prin:

$$\left| \frac{\mu_{01}}{w} \cdot \frac{1}{w \cdot h} \right| \quad (4)$$

unde μ_{01} este momentul statistic centrat de ordinul 0+1.

Prin experimentele efectuate s-au observat câteva proprietăți interesante ale factorilor individuali din care este compusă această metrică. S-a alcătuit o bază de date formate din imagini găsite ca urmare a căutării unor persoane foarte cunoscute, utilizând un motor de căutare pe web. S-au identificat manual doar pozele în care este realmente posibilă recunoașterea persoanei căutate. S-au comparat apoi, aceste rezultate, cu cele obținute cu ajutorul unui clasificator Adaboost, cu fiecare dintre cei patru factori calculați individual pe fiecare imagine. S-a obținut o rată de 87% clasificare, rezultat ce constituia de fapt principalul scop al cercetării. Dar, când se analizează în detaliu ponderile date fiecărui factor, în procesul de decizie, s-a observat că factorii FR și FO au fost ambii implicați în proporție de 31%, FS a fost implicat în proporție de 21% și factorul RSP a fost implicat în proporție de 17%. Cu alte cuvinte, programul de recunoaștere automată este mai mult influențat de orientarea feței și de trăsăturile de raport, dublu decât este influențat de culoarea pielii din imagine.

Deoarece sistemul supervizat a fost antrenat cu imagini selectate manual referitor la proprietatea lor de a permite identificarea persoanei reprezentate (înfățișate), putem trage câteva concluzii despre proprietățile cognitive care intervin preponderant în procesul de recunoaștere a feței umane.

Din păcate, în această etapă, trebuie să luăm în considerare aceste aspecte, cu precauție. Unele influențe asupra rezultatelor obținute, pot proveni din diverse cauze cum ar fi de exemplu exact opțiunea de a alege poze ale unor persoane celebre, artiști, care apar uneori cu machiaj și coafuri minuțios realizate, cu poze în condiții speciale de iluminare. Acest protocol a ridicat câteva aspecte interesante și necesită

experimente mai aprofundate, pe un număr mai mare de poze, utilizând mai multe imagini cu „persoane clasice”, bine cunoscute, pentru concluzii mai robuste.

Trăsăturile utilizate pentru diarizarea fețelor (unde termenul de diarizare se referă la identificarea fețelor într-o secvență video și adnotarea lor), pot fi descriptorii locali. Putem utiliza diverse alte abordări, cum ar fi modelele active de detectare a aparițiilor, dar principala limitare pentru astfel de procese rezidă în faptul că, în înregistrările video, fețele au în general o mică rezoluție video. Când avem la dispoziție doar câteva sute de pixeli pentru identificarea feței, acești algoritmi nu mai prezintă o foarte bună fiabilitate. Pentru a îmbunătăți rezultatele, poate fi utilă integrarea unor informații externe în legătură cu felul în care persoana este îmbrăcată, precum și elementele care apar pe fundal, în spatele ei³¹.



Fig. 4. Reconstrucția imaginii „Histogram Backprojection” plecând de la histograma de culoare rezultată după aplicarea metodei de deplasare a mediei (MeanShift) pentru localizarea unor posibile nou ocurențe a unei persoane anterior detectate

Experimente realizate cu o metodă de rezumare video automată au arătat că trăsăturile referitoare la îmbrăcăminte sunt indicii foarte semnificative de identificare a cadrelor cheie care aparțin unor anumite selecții de secvențe video. Odată ce o față a fost detectată, regiunea localizată sub zona delimitată pentru cap, poate fi caracterizată cu trăsături de nivel scăzut (culoare și textură). De fiecare dată când o nouă figură umană este detectată, regiunea corespunzătoare zonei de îmbrăcăminte poate fi comparată cu descriptorii precedent identificați în înregistrare. În cazul unei bune potriviri, aparițiile celor două fețe pot fi contopite într-un singur cluster, sau un instrument mai sofisticat de recunoaștere a feței poate fi aplicat pentru a confirma (sau nu) că cele două fețe corespund sau nu, aceleiași persoane.

Descrierea costumelor poate fi, de asemenea, o aplicație în sine, constituind o modalitate de a îmbunătăți instrumentele soft de detecție a fețelor umane. Este bine cunoscut faptul că astfel de instrumente pot da greș dacă unele condiții referitoare la orientare, iluminare, centrare, rezoluție etc. nu sunt îndeplinite. Când o persoană a

³¹ Gaël Jaffré, Philippe Joly, *Improvement of a Person Labeling Method Using Extracted Knowledge on Costume*, in 11th Int. Conf. on Computer Analysis of Images and Patterns (CAIP 2005), Versailles, France, 5-8/09/2005, A Gagalowicz, W Philips (Eds.), Springer-Verlag LNCS 3691, pp. 489-497, sept. 2005.

fost detectată și descriptorii de îmbrăcăminte corespunzători au fost stocați, un nou cadru (un frame) care urmează, poate fi analizat și sub aspectul găsirii unor regiuni asociate cu acești descriptorii. Dacă o astfel de regiune poate fi găsită și dacă aceasta verifică un set minimal de condiții (formă, mărime, conectivitate etc.), se poate decide că persoana corespunzătoare apare în cadru, chiar dacă fața nu a fost detectată (v. Fig. 4).

Pentru clusterizarea aparițiilor unor persoane, pot fi utilizați descriptorii locali și de culoare a pielii, dar și culoarea costumului sau descriptorii de textură. Tabelul următor dă indicii referitori la contribuția individuală și rezultatul scontat când în timpul procesului de clusterizare se aplică o fuziune timpurie de modalități de analiză.

Tabelul 1. Rata erorilor de clusterizare („cluster error rate”) obținută prin utilizarea diversilor descriptorii pe înregistrări tipice TV³².

	SIFT	Culoarea pielii (Skin color)	Culoare costum, îmbrăcăminte (Costume Color)	Textură costum, îmbrăcăminte (Costume Texture)	Fuziune (Fusion)
CER (%)	28.4	56.6	16.8	55.5	13.0

Rata de eroare a clusterizării dată mai sus, ia în calcul cea mai bună potrivire între clusterii de referință și clusterii automat generați de sistem, și durata conținutului procesat:

$$CER = \frac{\sum_{All\ shots} dur(shot), (\min(Nr(shot), Ns(shot)) - Nc(shot))}{\sum_{All\ shots} dur(shot), Nr(shot)} \quad (5)$$

unde, pentru un shot dat, *dur* este durata, *Nr* și *Ns* reprezintă numărul persoanelor identificate în referință și respectiv returnate de sistem, *Nc* este numărul de potriviri corecte pentru acel shot, cuprins între referință și clusterile sistemului.

Ratele de recunoaștere corectă variază în funcție de conținutul procesat. Rata de erori în clasificarea realizată cu ajutorul metodei fuziunii trăsăturilor este cuprinsă între 50% în seriale și aproximativ 9% pentru talk-show-uri tipice. S-ar putea crede că mixarea clusterizării audio și video ar putea conduce la rezultate mai bune decât fiecare din abordările individuale (fie audio, fie video). Dar, rezultatul este un pic diferit. Când se mixează elementele audio și video, am dori să asociem vocea unui personaj cu figura acestuia (în cazul nostru, mai precis, cu fața și îmbrăcămintea, costumul). Dar, în conținutul audiovizual, fața de pe ecran nu corespunde întotdeauna cu vocea de pe înregistrarea sonoră corespunzătoare filmării. Adnotarea automată este mult mai dificilă atunci când dorim să avem o corespondență voce-figură a aceleiași persoane în indexare (vocea și fața persoanei să aparțină aceluiași cluster),

³² Elie El Khoury, *Unsupervised Video Indexing based on Audiovisual Characterization of Persons*, PHD Thesis, University Paul Sabatier, June 2010.

O ipoteză comună în procesul de clusterizare constă în a considera că vocea și figura care apar cel mai adesea simultan, aparțin aceleiași persoane. Pe baza acestei presupunerii, a modalitate simplă de a fuziona clusterelor separat obținute din benzilor audio și video ale înregistrării, constă în a calcula mai întâi matricea de coocurențe.

$$M = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1nf} \\ m_{21} & m_{22} & \dots & m_{2nf} \\ \dots & \dots & \dots & \dots \\ m_{m1} & m_{m2} & \dots & m_{nmf} \end{bmatrix} \quad (6)$$

unde n_v este numărul de cluster vocale, n_f este numărul de cluster vizuale (utilizând descriptorii pentru figură și costume) și m_{ij} este numărul care arată de câte ori vocea i^{th} poate fi auzită atunci când privim fața persoanei j^{th} din înregistrare.

Înainte de a căuta maximul co-ocurenței este necesară normalizarea. Doi pași sunt obligatorii. Primul constă în calculul densității de probabilitate pentru fiecare figură condiționată de voce și respectiv calculul densității de probabilitate pentru fiecare voce condiționată de apariția simultană a figurii. Aceasta conduce la generarea a două noi versiuni ale lui M normalizate (a) rând cu rând și respectiv (b) coloană după coloană. Cele două noi matrice pot fi fuzionate utilizând un operator de fuziune cum este max, mean, sau product, între coeficienții aflați în aceeași locație. În matricea care rezultă, cu cât este mai mare coeficientul m_{ij} , cu atât este mai probabil faptul că vocea i și figura j aparțin aceleiași persoane.

Rezultatele obținute printr-un astfel de proces pot fi îmbunătățite dacă se ia în considerație că, de exemplu, modul în care este îmbrăcată o persoană este probabil să se schimbe pe parcursul unui film. În acest caz, ideea este de a genera într-o primă etapă cluster mic de voci și figuri cât mai precis posibil (cât mai pure, cât mai clare) și apoi să se fuzioneze clusterelor vizuale care pot fi asociate aceleiași voci luând în considerare valorile din matricea de co-ocurență. Două cluster vocale asociate acelui cluster vizual pot fi fuzionate în același mod. Procedând astfel, pe programe TV, s-au putut obține o rată de eroare în clusterizare CER de aproximativ 17%.

Identificarea persoanei într-o înregistrare audio-vizuală este încă o problemă deschisă pentru cercetare. De exemplu, mărimea feței ridică unele probleme încă nerezolvate. De câți pixeli avem nevoie pentru a fi capabili să identificăm o persoană? Câte persoane pot fi identificate într-o scenă aglomerată? Altă direcție care poate fi abordată în acest domeniu constă în utilizarea imaginilor de pe web ca o sursă de cunoaștere pentru identificarea persoanelor (cel puțin pentru identificarea persoanelor celebre).

Aspectele cognitive pot constitui o sursă de inspirație în rezolvarea unor probleme din acest domeniu. Ne putem focaliza pe comportamentul uman pentru a extrage câțiva descriptori sincretici. Am putea să emulăm modul în care memoria este implicată în procesul de identificare pentru a determina ce fel de trăsătură trebuie să utilizăm în strategiile de recunoaștere „short-term / long-term” (pe termen scurt – termen lung), prin crearea de modele. În final, știm că o limitare drastică a

instrumentelor de identificare automată este datorată inabilității acestora de a lua în calcul variațiile de voce, sau de aspect, ale unei persoane. Am putea studia și lua în considerație ce tip de seturi de trăsături sunt invariante în timp, sau să încercăm să modelăm felul în care acestea evoluează pentru a permite identificarea în condiții de variabilitate.

5. Analiza structurii

Crearea unui film urmează multe reguli privind diverse aspecte ale procesului de producție: felul în care aparatul de filmat trebuie să se miște pentru a efectua o anumită acțiune; modul în care muzica și efectele sonore trebuie amestecate cu conținutul vorbit de pe coloana sonoră; numărul de shot-uri în care trebuie împărțită o scenă pentru cea mai bună ilustrare a conținutului. Aceste reguli sunt predate în școlile de producție audiovizuală, precum și în cărțile dedicate acestui domeniu. Deși unele dintre acestea sunt destul de evidente, trebuie să luăm în considerație că nu toți consumatorii de conținut audiovizual au beneficiat de o educație academică legată de vizionarea televiziunii (spre deosebire de citirea conținuturilor textuale). De exemplu, capacitatea empirică de a privi un film nu este atât de evidentă dacă luăm în considerare modul în care au fost percepute primele filme din istoria cinematografului. Putem găsi numeroase relatări (probabil exagerate) despre felul în care privitorii ieșeau fugind din sala de cinema când vedeau, pe ecran, trenul intrând în gară în filmul „L'arrivée d'un train en gare de la Ciotat”. Chiar și în ziua de astăzi este nevoie de timp pentru a aminti publicului faptul că personajele nu reprezintă persoane reale.

Scrierea și editarea unui film reprezintă activități limitate de capacitatea publicului de a-i înțelege conținutul, precum și de unele caracteristici de producție.

Pentru ușurarea acestor procese au fost întocmite o serie de gramatici audiovizuale mai mult sau mai puțin formale³³. În astfel de gramatici aflăm că, pentru a extinde cronologia temporală a unei scene, se pot folosi efecte de tranziție treptată.

Muzica ar trebui să înceapă la sfârșitul unui capitol și înainte de începutul următorului. Aceste reguli pot fi folosite în mod direct pentru a construi un instrument de analiză automată³⁴.

Această proprietate poate fi aplicată și în cazul structurării evenimentelor sportive. Pentru acest tip de conținut, observăm un cadru de producție extrem restrictiv, care permite identificarea acțiunilor grație unei analize pe bază de HMM. În aceste cazuri, arhitectura HMM poate fi inspirată în mod aprioric chiar de normele sportive³⁵.

³³ Daniel Arijon, *Grammar of the film language*, Silman-James Press, 1991.

³⁴ Philippe Aigrain, Philippe Joly, Véronique Longueville, *Medium-Knowledge-Based Macrosegmentation of Video into Sequences*, in *Intelligent Multimedia Information Retrieval*, Mark T. Maybury (Eds.), AAAI Press and MIT Press, MIT Press, Cambridge, Massachusetts, pp. 159-173, 1997.

³⁵ Ewa Kijak, Guillaume Gravier, Lionel Oisel, Patrick Gros, *Audiovisual Integration for Sport Broadcast Structuring*, *Multimedia Tools and Applications*, 30:289-312, 2006.

Regulile de producție sunt, în general, induse de unele aspecte practice (pentru a arăta scene care se petrec în locații diferite, sau în momente diferite, este nevoie ca ele să fie tăiate în shot-uri). Sunt implicate, de asemenea, și reguli sintactice pentru a fi generate efecte de ritm sau pentru a se intensifica linia dramatică (un flashback, de exemplu).

S-a văzut, într-un paragraf anterior, că o aplicație alternativă a analizei similarității ar putea fi constituită pe ideea de a identifica structura programului³⁶. Am dorit să ne focalizăm pe modul în care evenimentele sunt temporal organizate în cadrul conținutului, pentru a caracteriza faptul că acele reguli de producție pot fi aplicate, în mod intenționat, pentru a diferenția un anumit document audiovizual de altele din aceeași categorie.

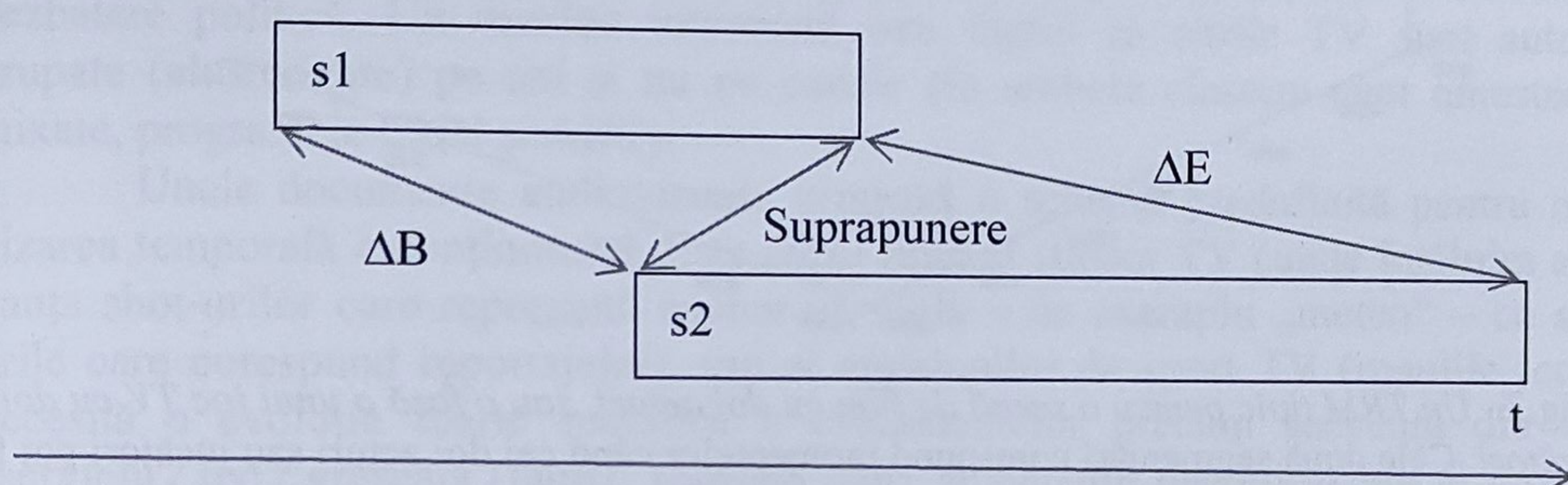


Fig. 5. Cei trei parametri utilizați pentru caracterizarea relației temporale dintre două segmente: diferența dintre pozițiile de început (ΔB), pozițiile de final (ΔE), precum și sfârșitul primului și începutul celui de al doilea segment (Suprapunere - Lap)

Astfel, am propus o metodă generală și nesupervizată bazată pe relațiile temporale dintre evenimente³⁷. Intrarea o reprezintă rezultatul diverselor procese de segmentare temporală aplicate aceluiași document audiovizual. Aceste segmentări pot viza identificarea segmentelor în care se pot auzi dialoguri sau aplauze, sau, spre exemplu, segmentele unde pot fi văzute figuri sau mișcare. Se presupune că fiecare proces de segmentare generează un set de segmente care nu se suprapun și care nu sunt în mod necesar conectate. Dacă se iau în considerare două segmente, s1 și s2, generate de două procese de segmentare diferite, se pot calcula următorii trei parametri pentru a caracteriza relațiile lor temporale:

Acești parametri sunt calculați pentru fiecare două segmente înăuntrul ferestrei temporale (mărimea ferestrei depinde de mărimea segmentelor și a evenimentelor ce trebuie caracterizate). Fiecare relație temporală este apoi considerată ca reprezentând un punct în spațiul 3D.

Se calculează, mai apoi, histograma 3D a tuturor relațiilor observate dintre rezultatele celor două segmentări temporale ale aceluiași document. Am numit această histogramă Temporal Relation Matrix (TRM) – Matricea relațiilor temporale. Deoarece unui conținut îi pot fi aplicate numeroase și variate procese de segmentare,

³⁶ Siba Haidar, Philippe Joly, Bilal Chebaro, *Mining for Video Production Invariants to Measure Style Similarity*, in International Journal of Intelligent Systems, Wiley, vol. 21, nr. 7, pp. 747-763, July 2006.

³⁷ Zein Al Abidin Ibrahim, Isabelle Ferrané, Philippe Joly, *A Similarity-Based Approach for Audiovisual Document Classification Using Temporal Relation Analysis*, in EURASIP Journal on Image and Video Processing, EURASIP, vol. 2011, (on line), March 2011.

pot fi calculate multe TRM-uri. S-a demonstrat că o algebră poate fi pe complet definită pe aceste matrice. Un set de operatori, și un element neutru se pot defini pentru a deduce valoarea unui nou TRM, date fiind două valori cunoscute, sub un set minimal de condiții. Putem observa, moment cu moment, într-un subset al acestor TRM unele distribuții specifice, care pot fi folosite imediat pentru identificarea unui gen de program TV sau al unui rol jucat de un personaj într-un context dat.

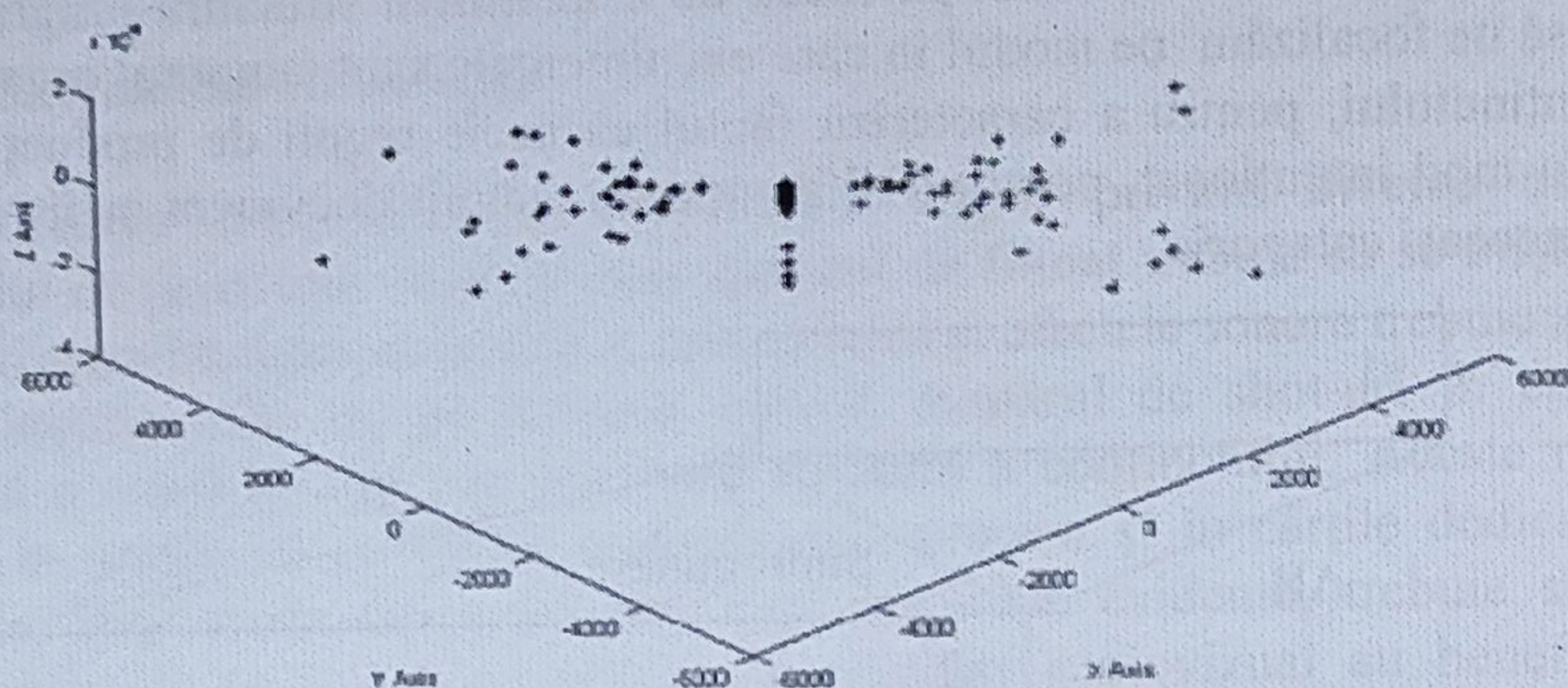


Fig. 6. Un TRM tipic pentru o scenă de film cu doi actori, sau o fază a unui joc TV cu doi jucători. Cele două segmentări corespund momentelor când cei doi actori sau jucători pot fi văzuți sau auziți. Punctele aliniate vertical corepund shot-urilor în care cele două persoane apar simultan. În partea stângă și în partea dreaptă, punctele reprezintă intervenția alternativă a acestor persoane.

Aceste distribuții pot fi analizate printr-un algoritm de cuantificare. Pot fi folosiți mai mulți algoritmi, cum ar fi K-means, sau scheme predefinite, cum ar fi cele din relația dată de James Allen³⁸, cunoscută drept relația Allen. Odată ce acest pas de cuantificare a fost aplicat, TRM-urile pot fi ele însele reprezentate printr-un mic vector de descriere pentru a identifica numărul relațiilor temporale.

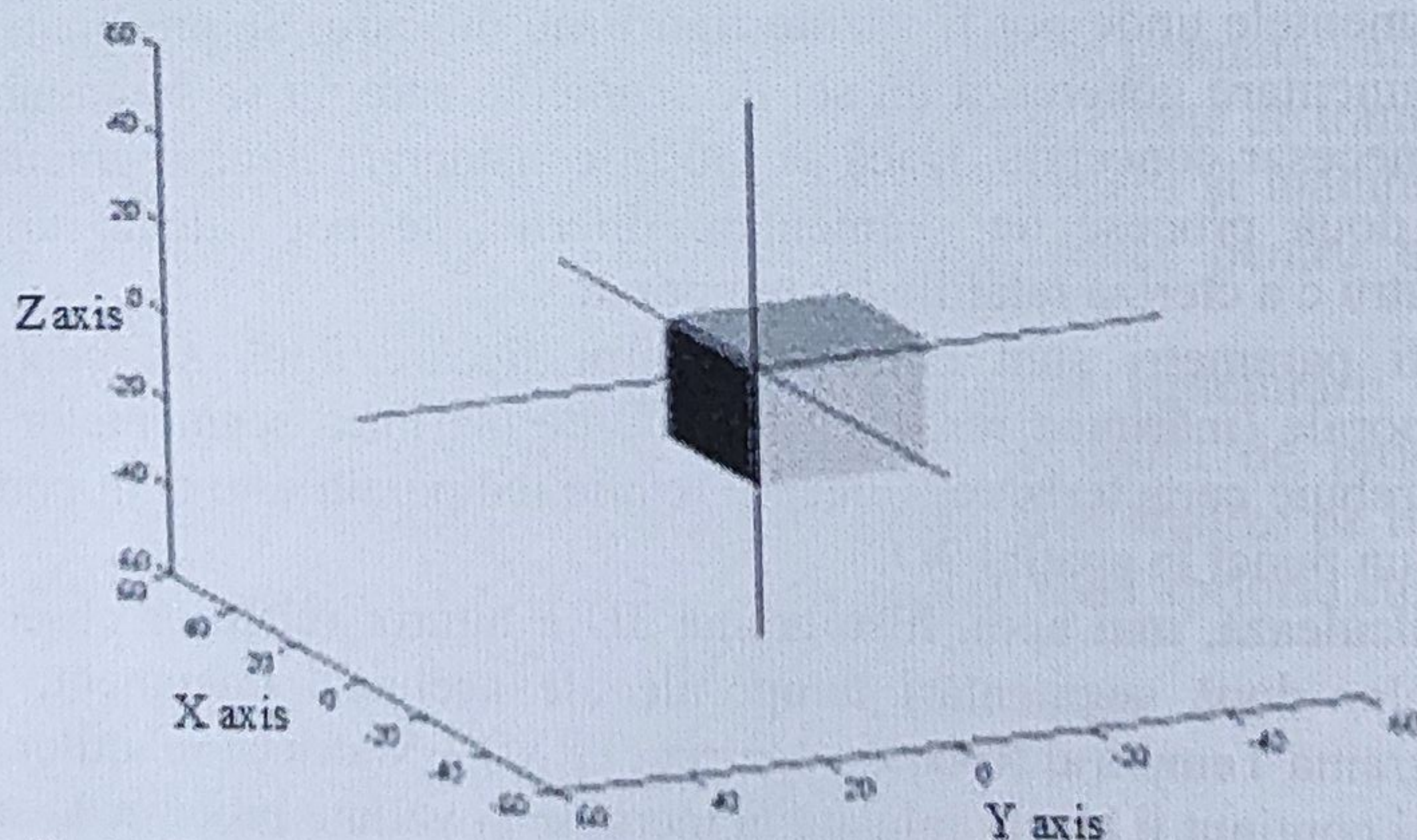


Fig. 7. Trei subspații, (trei plane) care corespund relațiilor Allen „se întâlnesc” (gri intermediar), „se suprapun” (gri deschis), „încep” (gri închis).

³⁸ James F. Allen, *Maintaining knowledge about temporal intervals*, in *Communications of the ACM* 26 Nov. 1983. ACM Press. pp. 832–843.

Folosind histograme normalizate calculate pe TRM cuantificate, putem, mai apoi, compara aceste matrice. Aplicând un simplu algoritm de clusterizare, am arătat în mod experimental că, clasificarea în gen a fost un simplu posibil output al acestui proces, și a fost posibilă obținerea unor rezultate interesante chiar și în clasificarea de tip sub-gen.

În Fig. 8, se observă cum au fost identificate 9 clustere (reprezentate pe axa verticală). Videoclipurile de la 1 la 12 reprezintă știri TV dintr-un set de conținuturi din anul 2004 (asociate, în majoritatea lor, cu clusterul 1), documentele de la 13 la 42 sunt știri TV dintr-un set de conținuturi din anul 2003 (clusterul 2). Documentele de la 43 la 49 sunt programe de football. Documentele de la 50 la 53 sunt filme. Documentele 54 și 55 sunt două jocuri TV diferite iar documentul 56 este o dezbatere politică. Un rezultat interesant este faptul că știrile TV sunt automat grupate (clusterizate) pe ani și nu pe canale (în ambele clustere sunt amestecate, mixate, programele CNN și ABC).

Unele documente audiovizuale urmează o schemă predefinită pentru organizarea temporală a conținutului. Este cazul tipic al știrilor TV (unde întâlnim alternanța shot-urilor care reprezintă anchor-ul, sigla – de exemplu „meteo” – cu shot-urile care corespund reportajelor), sau al emisiunilor de sport TV (regulile jocului necesită o evoluție foarte specifică a evenimentelor precum secvența din tenis „serviciu / joc / greșeală (fault)” repetată până la sfârșitul meciului) sau al jocurilor TV. Pentru acele documente, au existat multe propuneri de identificare a structurii folosind gramatici predefinite corespunzând, spre exemplu, HMM-urilor antrenate, sau, în cel mai bun caz, direct pe unele trăsături discriminante dedicate.

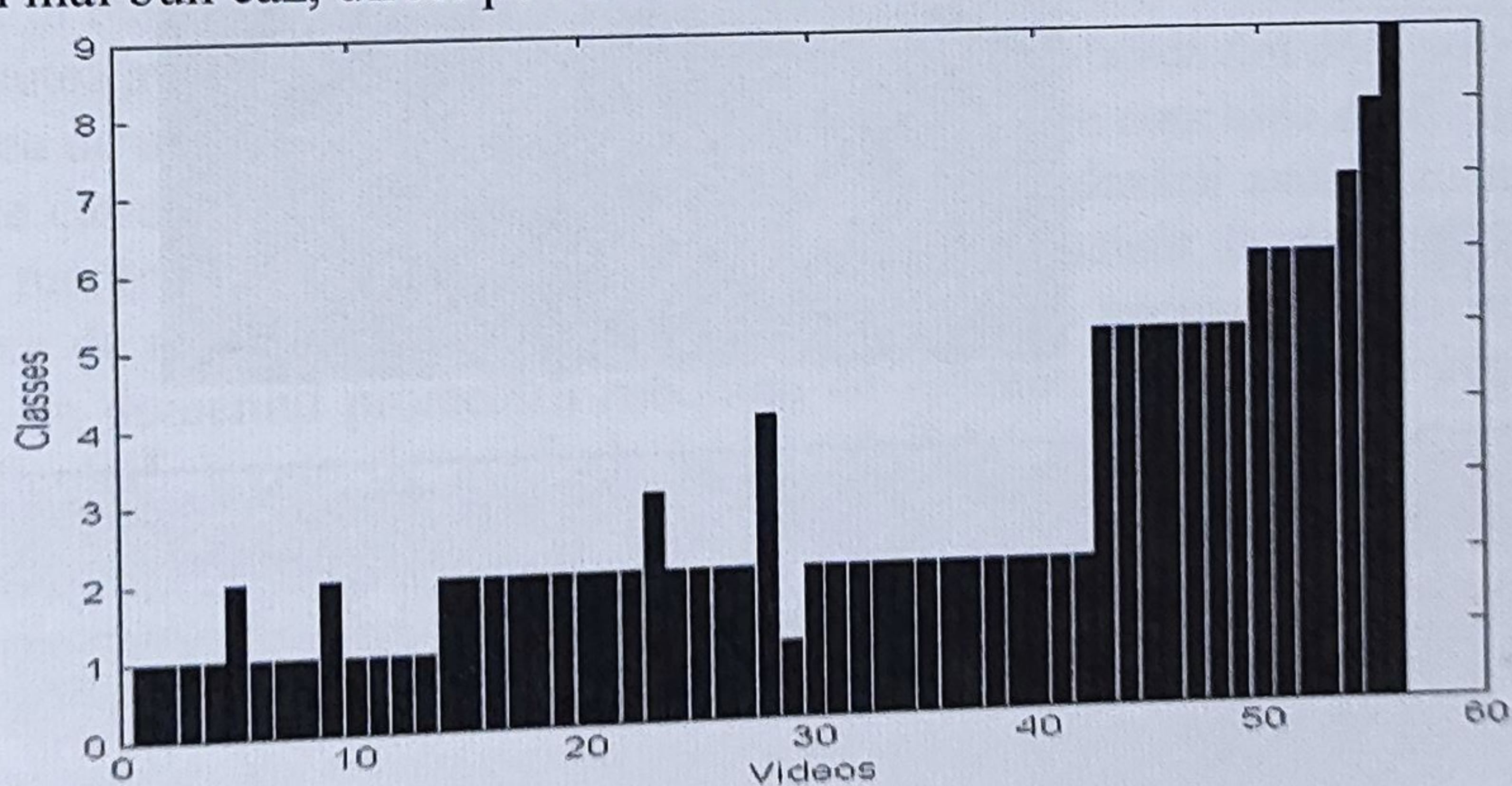


Fig. 8. Rezultat automat al clusterizării de documente video.

Pentru analiza automată a structurii temporale, cele mai dificile cazuri sunt identificate în filmele de ficțiune sau seriale. Pentru un astfel de proces, nici o gramatică nu poate fi identificată. Ne putem baza pe regulile de producție audiovizuală, dar, în acest caz, rezultatele pot fi calitativ-inferioare pentru conținuturi atipice. Putem folosi, de asemenea, abordări inspirate din teoria grafurilor pentru a încerca să adunăm shot-urile despre care se presupune că aparțin aceluiași episod. În unele lucrări recente, am propus combinarea unei abordări bazate pe graf cu rezultatele „diarizării” automate a personajelor (figură și voce) și un algoritm de clusterizare

pentru a propune un instrument de „video deinterlacing”³⁹. Aici, ideea este să se iau în considerație conținuturile mai multor evenimente despre care se presupune că au loc în același timp. Shot-urile sunt amestecate de-a lungul axei temporale cu scopul de a arăta o mică parte din fiecare eveniment la un moment dat. Pentru aceasta, căutăm seturi de personaje care apar cel mai mult în shot-urile succesive. Putem găsi apoi un set de clustere care corespund, în general, acelor evenimente filmate sau povestirilor și chiar mai mult, putem evidenția unele relații dintre aceste evenimente, sau relatări (un același personaj apare în două filmări, un același set etc.). Caracterizarea persoanelor și a modului lor de comunicare, într-o înregistrare audiovizuală, pot constitui de asemenea, date de intrare, de input, pentru identificarea rolului respectiv. Folosirea anumitor caracteristici legate de activitatea respectivă, de vorbire, durata de intervenție, energia semnalului, valoarea frecvenței fundamentale (pitch value) corepunzând unui singur personaj în cadrul unei ferestre temporale largi pot fi folosite într-un clasificator pentru identificarea rolurilor în programele de știri, precum siglă promo „anchor”, reporter, „journalist”, reporter ad-hoc „punctual journalist” etc. Pentru a avansa în direcția structurării temporale automate, putem identifica câteva scheme apriorice pentru ceea ce înseamnă un interviu între un jurnalist și relatarea unui martor ocular, de exemplu. Prin aplicarea unor scheme anterioare în zonele temporale unde două persoane interacționează, putem identifica unele scene specifice cum ar fi interviurile, prezentarea de informații, shot-urile de tranziție, pauzele etc.⁴⁰

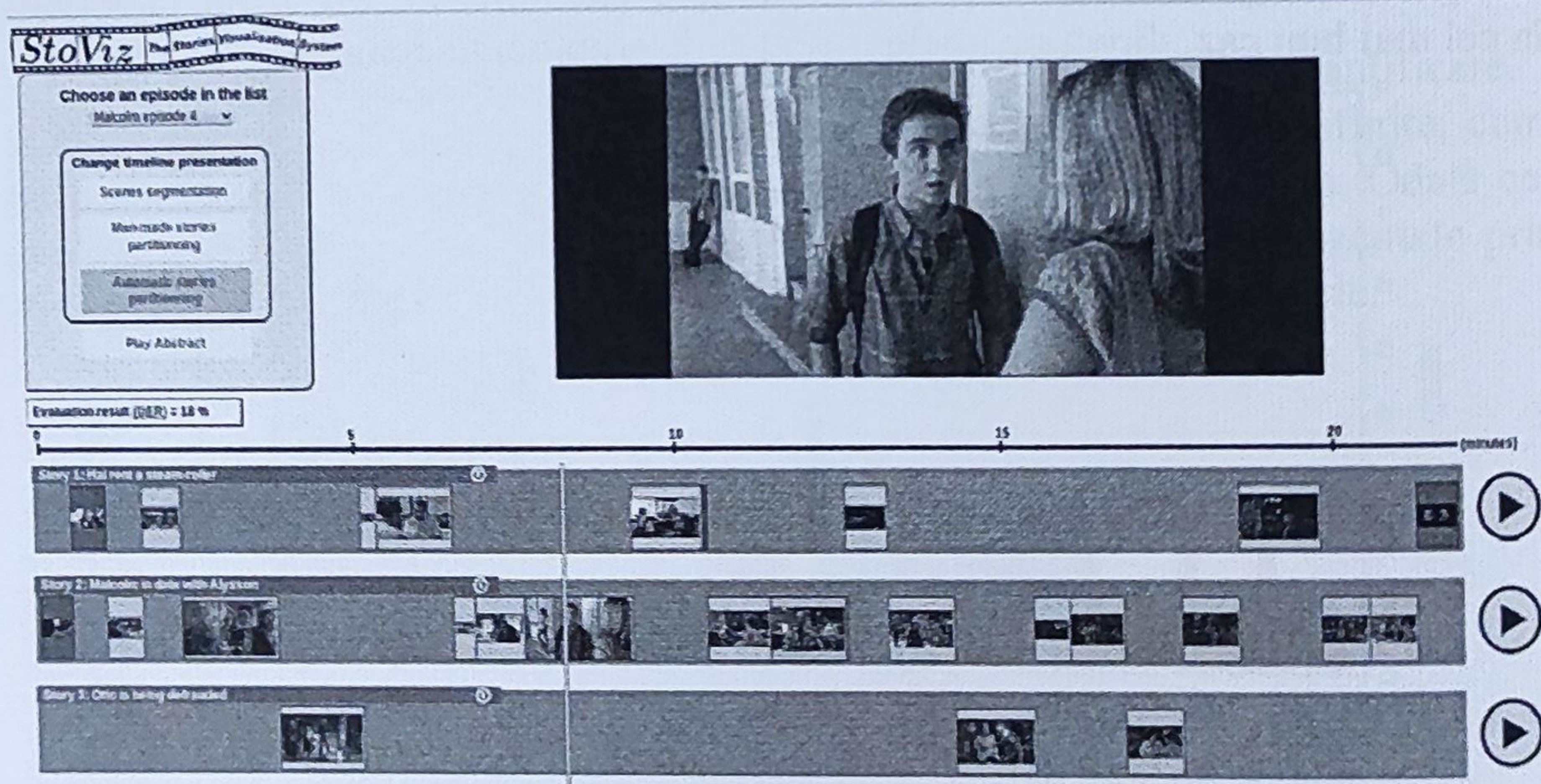


Fig. 9. Instrumentul de „video deinterlacing” pentru parcurgere (brows-are) automată a unei secvențe la un moment dat într-un videoclip, dezvoltat de către Ph. Ercolessi⁴¹.

³⁹ Interlacing (engl.) - interpătrundere (nota editorului), suprapunere parțială.

⁴⁰ Benjamin Bigot, Julien Piquier, Isabelle Ferrané, Régine André-Obrecht, *Detecting individual role using features extracted from speaker diarization results*, in *Multimedia Tools and Applications*, Springer-Verlag, vol. 60 nr. 2, pp. 347-369, Sept. 2012.

⁴¹ Philippe Ercolessi, Christine Senac, Philippe Joly, Hervé Bredin. *Segmenting Tv Series into Scenes Using Speaker Diarization*, in *Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2011)*, Delft - Pays bas, 13-15/04/2011, Delft University of Technology, (on line), apr. 2011.

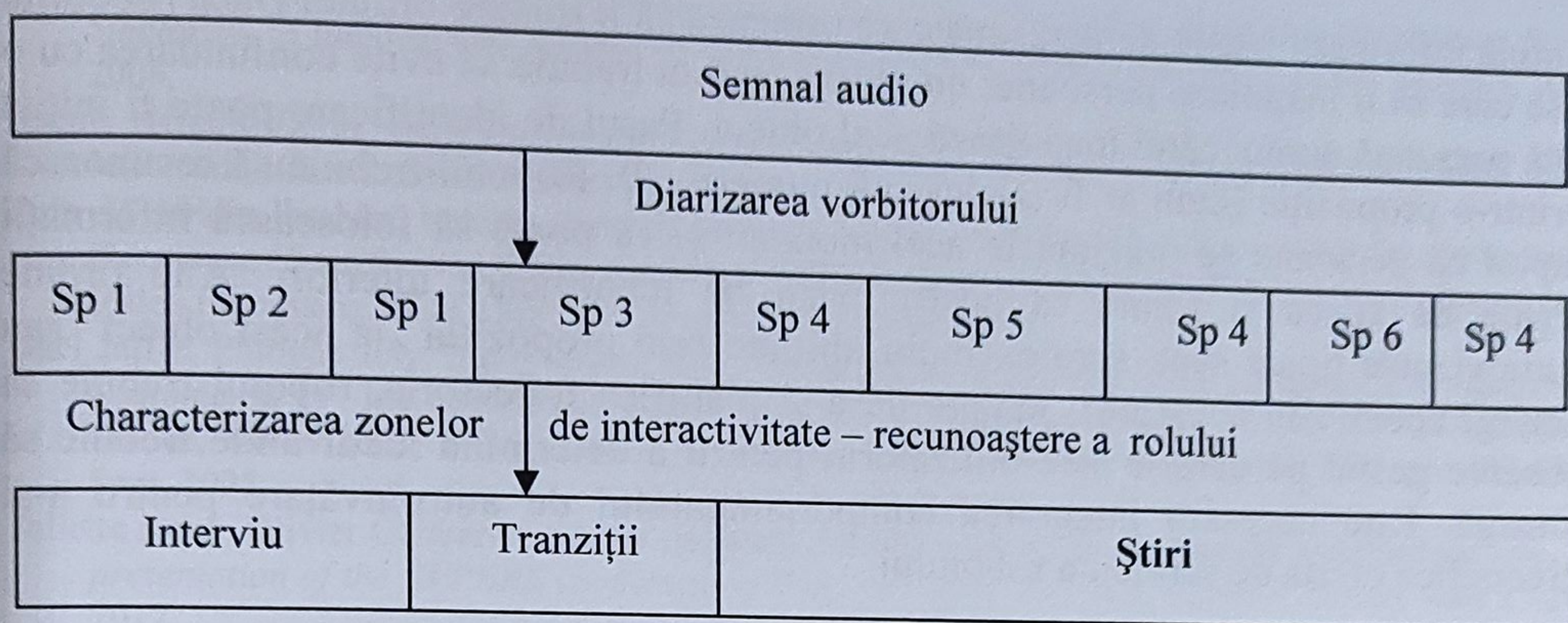


Fig. 10. Structurarea automată a programului de știri plecând de la diarizarea caracterului folosind rezultatele recunoașterii rolurilor.

Concluzii

Am prezentat câteva abordări dezvoltate pentru a trata problema indexării conținutului audiovizual. Primele realizări prezentate au constatat în extragerea caracteristicilor și estimarea similarităților semnăturilor construite pe baza acestor caracteristici. După cum am ilustrat în ultimele paragrafe, din ce în ce mai multe propuneri integrează, în prezent, parametri umani în procesul de caracterizare, precum și în etapa de calcul a distanței. Aceasta constituie o îmbunătățire obligatorie pentru a putea avansa în acest subiect, și pentru a putea aborda noi domenii, cum ar fi dezvoltarea de noi servicii de asistență la domiciliu. Vom conchide acest capitol cu o deschidere către acest domeniu de cercetare. Nevoia indexării audiovizuale în acest domeniu merge de la capacitatea de a urmări un pacient la domiciliu, până la dezvoltarea de noi dispozitive interactive care pot ajuta pacientul în activitățile sale cotidiene. În domeniul proiectării robotului de asistare, cercetătorii se confruntă cu multe dificultăți legate de auto-adaptarea robotului la mediul utilizatorului final. Problemele se pun, spre exemplu, în a găsi modalitatea de a-l face pe robot să recunoască numeroasele persoane familiare utilizatorului care locuiesc în aceeași casă, dar și sunetul telefonului (sau forma telefonului), care pot diferi de la o casă la alta. Prin urmare, unul dintre obiective este de a propune sisteme capabile să învețe cum arată elementele care prezintă interes. Un astfel de proces de învățare necesită existența unor atitudini independente, autonome, cum ar fi mișcarea către o țintă dată pentru a obține detalii mai clare legate de felul în care arată un obiect. Acest lucru necesită, de asemenea, capacitatea de a înțelege comenzi date atât în mod audio, cât și video⁴². Folosirea acestor două modalități reprezintă o manieră de a reduce erorile de interpretare, dar și de a conferi atribute care nu pot fi conferite într-o singură modalitate, chiar și în cazul unor comenzi simple, cum ar fi identificarea utilizatorului. Acest pas de identificare este necesar pentru a permite robotului să

⁴² Brice Burger, Isabelle Ferrané, Frédéric Lerasle, *Towards multimodal interface for interactive robots: challenges and robotic systems description*, in *Cutting Edge Robotics 2009*, KordicVedran (Eds.), IN-TECH, (on web) august 2009.

obțină informații legate de persoana care urmează să îi traseze ordine. Dacă robotului i se cere să îi înapoieze persoanei un obiect, acesta trebuie să evite confundarea cu o altă persoană atunci când înapoiează acel obiect. Pasul de identificare poate fi inițiat printr-o propoziție (cum ar fi „Salut, mă numesc...”). Robotul trebuie să recunoască faptul că persoana se prezintă în acel moment și că poate să folosească informații legate de vocea și figura sa pentru pașii de identificare ulteriori. Alte ordine audiovizuale tipice sunt, spre exemplu, inițiate prin propoziția „ia acest obiect” sau „mergi acolo”. În acest caz, înainte de a-și planifica traiectoria, robotul trebuie să observe gestul pe care îl face utilizatorul pentru a determina locul unde trebuie să meargă. Este necesară integrarea comportamentului de auto-învățare pentru a-și diversifica oferta de servicii a robotului.

Bibliografie

- Philippe Aigrain, Philippe Joly, Véronique Longueville, *Medium-Knowledge-Based Macrosegmentation of Video into Sequences*, in *Intelligent Multimedia Information Retrieval*, Mark T. Maybury (Eds.), AAAI Press & MIT Press, MIT Press, Cambridge, Massachusetts, pp. 159-173, 1997.
- James F. Allen, *Maintaining knowledge about temporal intervals*, in *Communications of the ACM* 26 ACM Press. pp. 832-843, Nov. 1983.
- Régine André-Obrecht. *A new statistical approach for automatic speech segmentation*, in *Transactions on Audio, Speech & Signal Processing*, IEEE, vol. 36 nr. 1, pp. 29-40, 1988.
- Daniel Arijon, *Grammar of the film language*, Silman-James Press. 1991.
- Sunkyoung Baek, Miyoung Cho, Myungwon Hwang, Pankoo Kim. *KANSEI-Based Image Retrieval Associated with Color*, in *Fuzzy Logics and Applications. Lecture Notes in Computer Science*. volume 3849. pp. 236-333. 2006.
- Benjamin Bigot, Julien Pinquier, Isabelle Ferrané, Régine André-Obrecht, *Detecting individual role using features extracted from speaker diarization results*, in *Multimedia Tools and Applications*, Springer-Verlag, vol. 60 nr. 2, pp. 347-369, Sept. 2012.
- Hervé Le Borgne, Nathalie Guyader, Anne. Guérin-Dugué, Jeanny Hérault. *Classification of images: ICA Filters vs Human perception*. In *Proc. Of ISSPA 2003*
- Brice Burger, Isabelle Ferrané, Frédéric Lerasle, *Towards multimodal interface for interactive robots: challenges and robotic systems description*, in *Cutting Edge Robotics*, KordicVedran (Eds.), IN-TECH, (on web) august 2009.
- Vincent Charvillat, Wei Tsang Ooi, Romulus Grigoraș, Géraldine Morin, *Crowdsourced Automatic Zoom and Scroll for Video Retargeting* în *ACM Multimedia*, Florence, 25/11/2010-29/11/2010, ACM, (suport electronic), 2010.
- Alain de Cheveigné, Hideki Kawahara. *YIN, a fundamental frequency estimator for speech and music*, in *J. of Acoustical Society of America*, 111 (4). pp. 1917-1930. April 2002.
- Michel Chion, *L'Audio-Vision, son et image au cinéma*, Ed. Nathan Université, 1991.
- Matthew Cooper, and Jonathan Foote, *Summarizing Popular Music via Structural Similarity Analysis*, in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2003)*.
- Bertrand Delezoide, *Hierarchical film segmentation using audio and visual similarity*, in *Proceedings of the IEEE International Conf. on Multimedia & Expo (ICME'05)*, 2005.
- Philippe Ercolessi, Christine Senac, Philippe Joly, Hervé Bredin. *Segmenting Tv Series into Scenes Using Speaker Diarization*, in *Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2011)*, Delft - Pays bas, 13-15/04/2011, Delft University of Technology, (on line), apr. 2011.
- Mark Everingham, Andrew Zisserman. *Automatic Person Identification in Video*, in *Image and Video Retrieval*, Vol 3115, LNCS 2004.
- Patrice Guyot, Julien Pinquier, Régine André-Obrecht, *Water Flow Detection From a Wearable Device With a New Feature, The Spectral Cover* in *International Workshop on Content-Based Multimedia Indexing (CBMI 2012)*, Annecy, France, 27-29/06/2012, IEEE, pp. 139-142, June 2012.

Aspecte cognitive în indexarea automată a conținutului audiovizual

- Siba Haidar, Philippe Joly, Bilal Chebaro, *Mining for Video Production Invariants to Measure Style Similarity*, in International Journal of Intelligent Systems, Wiley, vol. 21 nr. 7, pp. 747-763, July 2006.
- Jeanny Hérault, *De la rétine aux circuits neuromorphiques*. Chap, 3. in „Les systèmes de vision” J.M. Jolion ed., IC2 col., Hermes, Paris, 2001.
- Zein Al Abidin Ibrahim, Isabelle Ferrané, Philippe Joly. *A Similarity-Based Approach for Audiovisual Document Classification Using Temporal Relation Analysis*, in EURASIP Journal on Image and Video Processing, EURASIP, vol. 2011, (on line), March 2011.
- Gaël Jaffré, Philippe Joly, *Improvement of a Person Labeling Method Using Extracted Knowledge on Costume*, in 11th Int. Conf. on Computer Analysis of Images and Patterns (CAIP 2005), Versailles, France, 5-8/09/2005, A Gagalowicz, W Philips (Eds.), Springer-Verlag LNCS 3691, pp. 489-497, Sept. 2005.
- Juliette Kahn, Olivier Galibert, Ludovic Quintard, Matthieu Carré, Aude Giraudel and Philippe Joly. *A presentation of the REPERE Challenge*, in Proc. of CBMI 2012. pp. 102-107. Annecy June 27-29, 2012.
- Daoudi Khalid, Jérôme Louradour (2009), *A Comparison between sequence kernels for SVM speaker verification*, 4241-4244. In Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP).
- Elie El Khoury, *Unsupervised Video Indexing based on Audiovisual Characterization of Persons*, PHD Thesis. University Paul Sabatier, June 2010.
- Elie El Khoury, Christine Senac, Julien Piquier, *Improved Speaker Diarization System for Meetings*, in IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009, Taipei, Taiwan, 19÷24/04/09, IEEE, pp. 4241-4244, 2009
- Ewa Kijak, Guillaume Gravier, Lionel Oisel, Patrick Gros, *Audiovisual Integration for Sport Broadcast Structuring*. Multimedia Tools and Applications, 30:289-312, 2006.
- Ikram Ben Kouas, Philippe Joly, *Automatic Keyface Selection for Known People Identification in Images*, in Proc. of ECMS 2011. IEEE Explore, 2011.
- B. S. Manjunath, Philippe Salembier, Thomas Sikora, *Introduction to Mpeg-7: Multimedia Content Description Interface*. Editor: John Wiley & Sons Ltd. 2002.
- Geoffroy Peeters, Xavier Rodet, *Signal-based Music Structure Discovery for Audio Summary Generation*. ICMC03 Singapore, Singapore, 29 sept. - 4 oct. 2003.
- Jeremy Philippeau, Julien Piquier, Philippe Joly, Jean Carrière. *Dynamic organization of audiovisual database using a user-defined similarity measure based on low-level features*, in IEEE International Conference on Image Processing (ICIP 2008), San Diego, California, U.S.A., 12/10/08-15/10/08, IEEE, pp. 33-36, 2008.
- Jean-Luc Rouas, Jérôme Farinas, François Pellegrino, Régine André-Obrecht, *Rhythmic unit extraction and modelling for automatic language identification*, in Speech Communication, Elsevier, Vol. 47 N. 4, pp. 436-456, 2005.
- Xavier Sevillano, Xavier Valero, Francesc Alias, *Audio and video cues for geo-tagging online videos in the absence of metadata*, in proc. of CBMI 2012. pp 217-222.
- Romain Tavenard, Laurent Amsaleg, Guillaume Gravier. *Model-based similarity estimation of multidimensional temporal sequences*, in Annales des Télécommunications 64, 5 (2009) 381-390.
- Simon Thorpe, Arnaud Delorme, Rufin VanRullen and Williams Paquier (2000) *Reverse engineering of the visual system using networks of spiking neurons*. IEEE International Symposium on Circuits and Systems 4 405-408.

Tehnologii de asistare a persoanelor cu deficiențe vizuale

CHRISTOPHE JOUFFRAIS¹

ABSTRACT. Visual disabilities constitute disadvantages with major impact on important areas of society, such as accessibility to information, to professional spheres or associations, access to culture and recreation. The present research focuses exclusively on spatial cognition in people with visual impairment, either directly or through interactive technologies, and visual compensation by these interactive technologies. Vision impairment, causes, consequences and the theory of spatial cognition for the design of interactive systems for people with visual impairments will be detailed in the idea of finding the optimal methods to compensate for those deficiencies.

KEYWORDS: *visual impaired persons, neurocognition, visual prosthesis.*

1. Introducere

Handicapul vizual consituie un dezavantaj cu un impact major în domenii importante ale societății, cum ar fi accesibilitatea la informații, la sferele sau asociațiile profesionale, accesul la cultură și la posibilitățile de recreere. Proiectul de cercetare descris, condus în cadrul Institutului de Informatică din Toulouse, Universitatea Paul Sabatier, se concentrează exclusiv asupra cogniției/cunoașterii spațiale la persoanele cu deficiențe vizuale, fie direct, fie prin intermediul tehnologiilor interactive. Se propun diverse tehnici de compensare a acestor deficiențe vizuale prin tehnologii interactive. Deficiențele de vedere, cauzele, consecințele lor și cadrul teoretic al cogniției spațiale pentru proiectarea de sisteme interactive destinate persoanelor cu deficiențe de vedere sunt detaliate în ideea găsirii metodelor optime de compensare a acestor deficiențe.

2. Contextul și cadrul teoretic al cercetării

Cercetările prezentate în acest capitol se centrează pe cogniția spațială în cazul persoanelor cu deficiențe vizuale, pornind de la aspectele fundamentale legate

¹ Christophe JOUFFRAIS, IRIT-ELIPSE, CNRS & UPS, Toulouse (autor) – text adaptat de Mihaela Costin, Academia Română filiala Iași, în stagiul de cercetare post-doctorală efectuat la IRIT, Institute of Research in Informatics, Toulouse, France, în cadrul proiectului „Societatea Bazată pe Cunoaștere – cercetări, dezbateri, perspective”, cofinanțat de Uniunea Europeană și Guvernul României, Fondul Social European, Programul Operațional Sectorial Dezvoltarea Resurselor Umane POSDRU 56815.

de percepție, până la aspectele aplicate în proiectarea și evaluarea dispozitivelor de înlocuire a capacităților senzoriale vizuale prin alte alternative senzoriale. Studiul se bazează pe disciplinele complementare, de la neuroștiințe și psihologie experimentală (care permit să înțelegem capacitățile și procesele senzoriale și cognitive însoțind deficiențele vizuale) către proiectarea tehnologiilor interactive augmentative, a interfețelor grafice pentru utilizatori (GUI), în studiul interacțiunii om-mașină pentru evaluarea impactului asupra cogniției spațiale a persoanelor cu deficiențe vizuale.

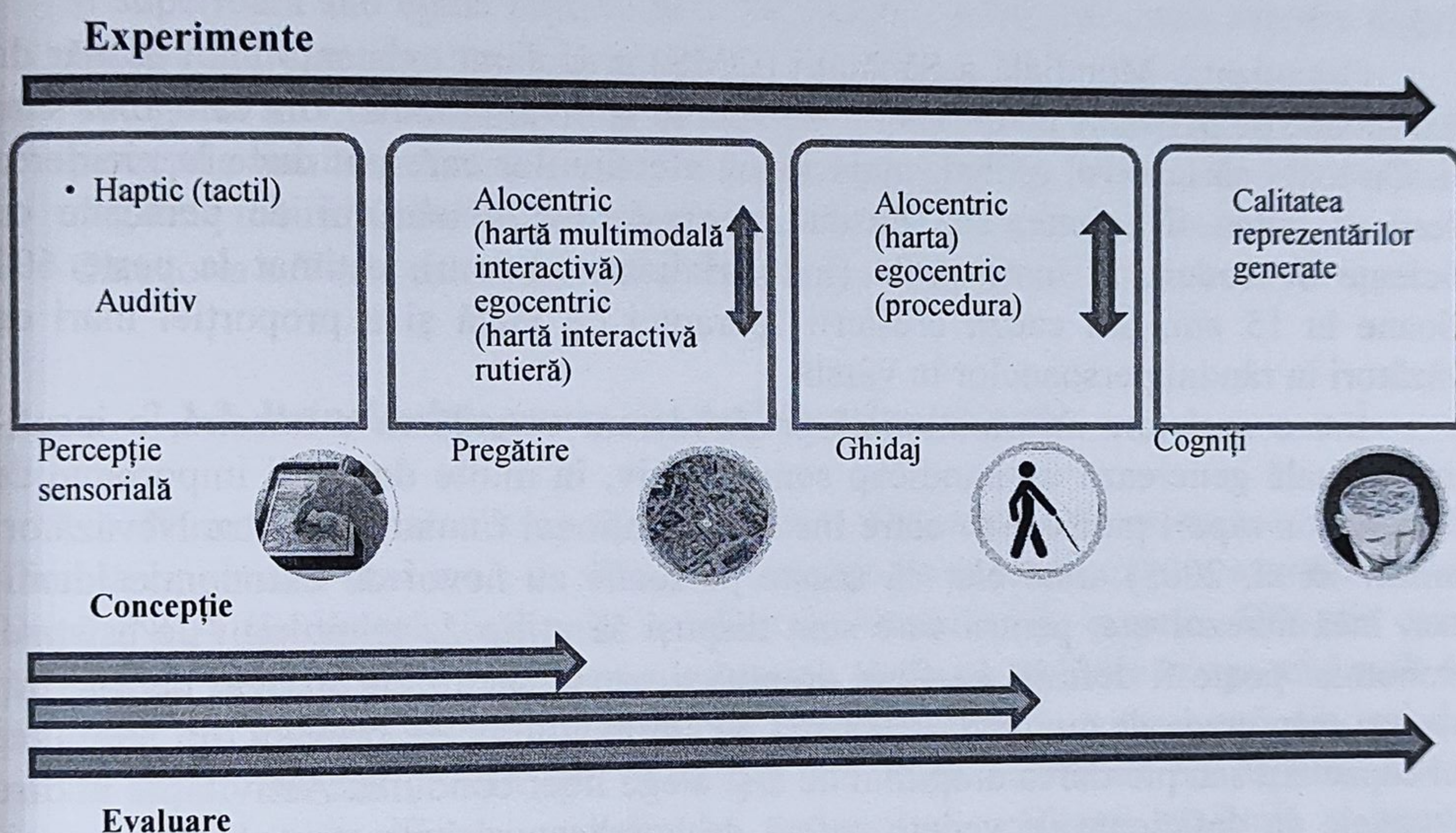


Fig. 1. Schema generală a proiectului: ansamblul etapelor de cercetare pentru realizarea tehnologiilor de asistare

Figura 1 reprezintă o schemă de ansamblu a etapelor de cercetare necesare pentru realizarea unor tehnologii de asistare a persoanelor cu deficiențe vizuale. Primul modul corespunde cercetărilor în domeniul psihologiei experimentale pentru percepția spațial-tactilă și auditivă directă sau mediată prin tehnologii de asistare. Următoarele două module ilustrează cercetările pentru tehnologiile de suplinire, de înlocuire, a văzului prin pregătirea unor modalități de facilitare a navigației GPS asistate vocal. Ultimul modul corespunde domeniului de cercetare în psihologia cognitivă privind calitatea reprezentării spațiale generate, folosind diverse tehnologii de asistare proiectată. Cercetările în psihologie influențează proiectarea de sisteme de asistare proiectată. Cercetările în psihologie pun în relație percepția și supleante, de substituție, iar cercetările fundamentale pun în relație percepția și cogniția (aspectele perceptive cu cele cognitive). Termenul alocentric exprimă o tendință de focalizare a interesului personal spre ceilalți, exterior (mai mult decât către sine, intenția personală fiind egocentrică)².

Este interesant de subliniat contextul social și cadrul teoretic în care se încadrează aceste cercetări. Studiile efectuate în domeniul științelor și tehnologiilor

² N. B. alocentrism – a nu se confunda cu altruismul, atitudine morală, care implică sacrificiul voluntar de sine, în timp ce orientarea alocentrică este pur psihologică.

cognitive pentru deficiențele de vedere se bazează pe studii de psihologie experimentală, care permit proiectarea unor dispozitive de supleanță (de suplinire) a funcțiilor vizuale. Fiecare prototip trece printr-o fază de evaluare, evidențiindu-se creșterea autonomiei persoanei și permite abordarea problemei cunoașterii spațiale atunci când este mediată prin tehnologii interactive.

2.1 Contextul mondial

Organizația Mondială a Sănătății (OMS) a declarat existența unui număr de 314 milioane de persoane cu deficiențe de vedere la nivel mondial, din care 15% sunt orbi. Cu toate că la nivel global, majoritatea afecțiunilor care pot duce la pierderea vederii ar putea fi evitate, se constată o creștere a numărului de persoane cu deficiențe de vedere în toate țările (industrializate sau nu), estimat la peste 600 milioane în 15 ani, din cauza creșterii speranței de viață și a proporției mari de nevăzători în rândul persoanelor în vârstă.

Într-o societate informațională preponderent vizuală ca a noastră, o insuficiență vizuală generează un handicap semnificativ, în multe domenii importante de activitate. Un raport publicat de către Institutul Național Canadian pentru Nevăzători (Simson³ et al. 2005) arată clar că aceste persoane au nevoi de autonomie identificate, încă nerezolvate, pentru care sunt dispuși să utilizeze tehnologii de asistare. Autonomia⁴ poate fi definită ca fiind dreptul și capacitatea unui individ de a-și trăi viața așa cum crede de cuviință. Pierderea de autonomie poate proveni din pierderea unei capacități sau pierderea dreptului de a-și alege liber conduita. Activitățile în care persoanele cu deficiențe de vedere suferă de handicapuri importante sunt comunicarea în scris și orientarea spațială (înțelegerea, perceperea mediului, a locației unui obiect, capacitatea de a circula, de a se dirija către o anumită țintă).

2.1.1 Deficiențele vizuale

Orbirea absolută este definită printr-o acuitate vizuală sub 1/100 după corecție sau când câmpul vizual este sub 20° – norma standard fiind de 180°. Acuitatea vizuală este raportul dintre distanța de la care un obiect este văzut de către subiect (numărătorul), și (numitorul) distanța de la care același obiect, fără a schimba caracteristicile sale, este văzut de către un subiect care nu are nici o deficiență vizuală. De exemplu, o acuitate de 6/60 semnifică faptul că obiectul observat la 60 de metri de către o persoană cu vederea normală, trebuie apropiat la 6 metri de persoana cu deficiență vizuală, pentru a fi perceput în același fel. Actualmente, OMS se află la noua revizuire a clasificărilor internaționale a maladiilor pentru reclasificarea deficiențelor vizuale luând în calcul acuitatea reziduală și câmpul vizual. Sunt definite cinci categorii de deficiențe vizuale, fie în funcție de scăderea acuității vizuale fie în funcție de diminuarea câmpului vizual:

³ Simson, Helen, Deborah Gold & Biljana Zuvella, *Inégalité des chances: Rapport sur les besoins des personnes aveugles ou handicapées visuelles vivant au Canada*, 2005.

⁴ Prin originea sa greacă, cuvântul „autonomie” semnifică „legea, regula, pe care și-o stabilește fiecare.”

- Categoria I-a: acuitate vizuală binoculară corectată inferioară raportului 3/10 și superioară sau egală raportului 1/10 cu un câmp vizual mai mic de 20°.
- Categoria a II-a: acuitate vizuală binoculară corectată inferioară raportului 1/10 și superioară sau egală raportului 1/20; practic, subiectul poate număra degetele mâinii de la 3 m.

Următoarele categorii corespund noțiunii de cecitate (a fi orb, nevăzător):

- Categoria a III-a: acuitate vizuală binoculară corectată inferioară raportului 1/210 și superioară sau egală raportului 1/50; practic, subiectul poate număra degetele mâinii de la 1m, dar nu de la 3m.
- Categoria a IV-a: acuitate vizuală binoculară corectată inferioară raportului 1/50; percepție luminoasă conservată, încă prezentă; practic, subiectul nu poate număra degetele mâinii de la 1m, iar câmpul vizual este inferior valorii de 5°.
- Categoria a V-a: cecitate/orbire absolută, lipsa percepției luminoase.

2.1.2 Principalele cauze ale deficiențelor vizuale

Numeroase maladii pot afecta vederea. Cataracta este prima cauză de orbire, fiind la originea a mai mult de 40% din cazurile de scădere acută a acuității vizuale. După INCA (Institutul Național Canadian de Nevăzători), în peste 50% din cazuri, persoanele au orbit ca urmare a unei cataracte. Această maladie atinge mai mult de o persoană din cinci după vârsta de 65 de ani, mai mult de o persoană din trei după vârsta de 75 de ani și aproape două din trei după vârsta de 85 de ani. Aceasta corespunde la opacifierea cristalinului provocând scăderea graduală a acuității vizuale până la orbire, dacă nu este tratată la timp. Tratamentul este chirurgical (extragerea cristalinului și implantarea unei lentile intra-oculare).

Glaucomul este a doua cauză de orbire, antrenând o presiune intraoculară care duce la compresia nervului optic, pentru care nu exista încă o metodă curativă. Degenerescenta maculară legată de vârstă, (DMLV) este principala cauză de orbire în țările industrializate. Aceasta corespunde la o atrofie a zonei centrale a retinei (zona țările industrializate. Aceasta corespunde la o atrofie a zonei centrale a retinei (zona centrală din fovea), rezultând în afectarea vederii centrale și provoacă în primul rând, disconfort la citit sau la condus putând evolua spre insuficiență vizuală severă. Frecvența sa crește cu vârsta, atingând 25% din persoane după vârsta de 80 de ani și este principala cauză de orbire la persoanele de peste 50 de ani. Nu există nici un tratament și cu îmbătrânirea populației, problema DMLV va crește cu siguranță.

Opacitatea corneei corespunde la o pierdere a transparenței corneei și poate apărea din diverse cauze cum ar fi fenomenele cicatriceale după traumatisme sau infecții corneene sau datorită dezorganizărilor structurale ale diverselor straturi corneene. Netratată, poate conduce la deficiențe de vedere semnificative.

Retinopatia diabetică este o consecință a leziunilor vaselor capilare de la nivelul retinei, datorită diabetului zaharat. În forma sa severă, proliferativă, poate conduce la insuficiență vizuală profundă. Aproximativ 10% dintre pacienți dezvoltă o deficiență vizuală gravă după 15 ani de diabet zaharat.

2.2 Cadrul teoretic

2.2.1 Cartografierea cognitivă

În proporții diferite, fiecare dintre noi avem conștiința spațială a mediului în care trăim. Datorită stimulărilor senzoriale în general (și informațiilor provenind de la organele noastre de simț), reușim să cunoaștem și să înțelegem diferitele locuri și situații. Aceste date recepționate senzorial, pe de o parte provin direct din mediu, datorită percepției directe a obiectelor care pot fi atinse, pipăite, mirosite, auzite, văzute (sau gustate), iar pe de altă parte provin din informațiile geografice parvenite prin intermediul altor surse, mai puțin directe cum sunt internetul, cărțile, televiziunea, radio, jurnalele, hărțile, machetele sau pur și simplu conversația. Integrarea acestor prime informații provenind din experiență și din învățare este un proces complex cunoscut sub numele de cartografiere cognitivă. Aceasta înglobează un ansamblu de operațiuni prin care individul obține, codează, stochează, își amintește și manipulează informația legată de mediul său spațial. Această informație acumulată corespunde atributelor, caracteristicilor și pozițiilor relative ale lucrurilor și persoanelor în mediu și este indispensabilă în procesul adaptiv de luare a unei decizii spațiale (v. Downs⁵ and Stea 1973).

Cunoașterea și înțelegerea mediului joacă un rol important în planificarea și efectuarea oricăror deplasări (Garling⁶ & Golledge 2001). Acestea permit de exemplu identificarea unui supermarket în apropiere, selecționarea celui mai bun drum pentru a putea ajunge la serviciu, sau decizia asupra locului în care am dori să ne petrecem vacanța. Aceste procese evoluează cu vârsta, experiența, învățarea, dar și cu apariția deficiențelor senzoriale. O hartă sau o reprezentare cognitivă (Tolman⁷ 1948) este deci o abstracție mentală (mai mult sau mai puțin corectă) a unui mediu, elaborată de un individ la un moment dat. Cartografierea cognitivă poate fi considerată ca un model intern al lumii în care trăim (Golledge⁸ & Stimson 1997). Una dintre caracteristicile reprezentărilor mentale spațiale este distincția dintre cele două tipuri de reprezentări denumite respectiv hartă sau drum.

Reprezentarea de tip hartă, maparea, este o reprezentare a unui ansamblu elaborat într-un cadru de referință alocentric, deci în care sistemul de coordonate este extern. Această reprezentare este bazată pe o cunoaștere a proprietăților topografice ale mediului, implicând localizarea obiectelor în raport cu un sistem de coordonate fixe. Este deci o reprezentare independentă de poziția individului. Rolul mapării este crucial în capacitatea persoanei de a determina configurația locurilor situate înafara câmpului său vizual, sau în stabilirea relațiilor spațiale între locuri care nu au fost

⁵ Downs, Roger M. & David Stea, *Image & environment: Cognitive mapping and spatial behavior*, Edward Arnold, London, 1973.

⁶ Garling, Tommy & Reginald G. Golledge, *Cognitive mapping and spatial decision-making. Cognitive mapping past, present and future*, ed. by R.M. Kitchin & S. Freundschuh, 44-65: Routledge, 2001.

⁷ Tolman, Edward C., *Cognitive maps in rats and men*, Psychological review 55.189, 1948.

⁸ Golledge, Reginald G. & Robert J. Stimson, *Analytical behavioural geography*, Guilford Press, 1997.

poate, încă, explorate fizic. Organizarea unei reprezentări de tip hartă este adeseori de tip ierarhic, deci spațiul este divizat în zone, fiecare zonă fiind descrisă succesiv.

Reprezentarea de tip hartă, contrar reprezentării de tip drum, permite o reorganizare a informațiilor spațiale (întoarceri, găsirea unor scurtături etc.).

Reprezentarea de tip drum este elaborată într-un cadru de referință ego-centric, deci într-un sistem de coordonate integrând reperele corpului, poziția celui care navighează. Cunoașterea spațiului se face plecând de la înregistrarea secvențială a punctelor de reper pe măsură ce individul navighează în acest mediu (de exemplu pe măsura parcurgerii unui traseu). Acest tip de cunoaștere este puțin plastică în sensul în care ea nu poate fi reorganizată (dacă un punct de reper a dispărut, sau dacă navigatorul a deviat, structura este întreruptă). Această tip de reprezentare pare a fi suficient pentru cazul în care deplasarea se realizează într-un mediu cunoscut, familiar, fără obstacole.

Simțul vizual este cel preponderent la om. Când se primesc simultan informații vizuale și informații senzoriale de alt tip, subiectul interpretează frecvent informațiile vizuale ca fiind corecte, chiar dacă acestea nu sunt. Acest fenomen a fost numit „captură vizuală” (Smyth⁹ & Wing, 1984) sau dominanță vizuală. Un exemplu pentru acest fenomen este conflictul vizual-tactil. Când un obiect este perceput vizual într-o anumită formă (oval, de exemplu), în timp ce el este perceput tactil sub o altă formă (de exemplu în forma literei C), percepția vizuală este favorizată (Smyth & Wing 1984). Deci uzual, percepția vizuală este cea care furnizează competențele și resursele de bază, permițând principalele interacțiuni cu mediul. Vederea permite conștientizarea propriei poziții și actualizarea informațiilor care provin din mediu în scopul navigării și construirii unei hărți.

O consecință a dominanței simțului vizual este că există diferențe în utilizarea celor două tipuri de reprezentare (hartă versus drum), pentru văzători și nevăzători, dar și între nevăzătorii congenitali și cei tardivi.

2.2.1.1 Navigarea

Termenul de navigare definește comportamentul de deplasare orientată către un anumit scop, cu toate procesele motorii, senzoriale și cognitive pe care le implică. Downs and Stea¹⁰ dau următoarea definiție: „procesul rezolvării unei clase de probleme spațiale, mișcarea unei persoane dintr-o locație pe suprafața pământului, în alta”. Conform acestui studiu, sunt departajate patru etape: orientarea în mediu, alegerea itinerariului, conservarea itinerariului bun și recunoașterea faptului că destinația este atinsă.

Prima etapă (orientarea) corespunde la cunoașterea locului unde se găsește persoana care navighează în raport cu alte locuri pe care aceasta în prealabil, le cunoștea deja. Această etapă de orientare, denumită și stabilirea direcției, permite subiectului să cunoască poziția în care se găsește și cum este orientat în raport cu mediul.

⁹ Mary M. Smyth, *Memory for movements*, in Mary M. Smyth Alan M. Wing (Eds.), *The psychology of human movement*, London: Academic Press, 1984.

¹⁰ Downs, Roger M. & David Stea, *Image & environment: Cognitive mapping and spatial behavior*, Edward Arnold, London, 1973.

Cea de a doua etapă (alegerea itinerariului) impune ca navigatorul să facă o legătură cognitivă între punctul de plecare (de exemplu poziția sa curentă) și destinația dorită, alegând un itinerariu care leagă cele două puncte. În numeroase medii este dificil să se atingă direct o destinație. Existența multiplelor obstacole (imobile, râuri) împiedică adeseori deplasarea în linie dreaptă. O rețea de posibile trasee permite calculul multiplelor itinerare alternative. Alegerea unui anumit traseu depinde de numeroși factori externi ca de exemplu distanța de parcurs, comparată, sau factori subiectivi ca sentimentul de securitate, simplitatea, stresul generat de anumite etape, sau punctele de interes care se găsesc pe traseu. Această alegere poate fi făcută înainte de plecarea pe traseu, ceea ce permite obținerea instrucțiunilor de navigație incluzând punctele de decizie și de orientare (Timpf¹¹ ș.a. 1992).

Cea de-a treia etapă, păstrarea itinerariului corect, constă în a se asigura că, în timpul deplasării, itinerariul ales nu a fost părăsit. Acest lucru implică ca cel care se deplasează să facă alegerile corecte la nivelul punctelor de decizie. Făcând acest lucru, călătorul își evaluează în mod constant orientarea în spațiu și se poziționează pe harta sa cognitivă. Reperele joacă un rol important în această sarcină și sunt verificate în mod regulat. În plus, deseori, navigatorul trebuie să iasă din itinerariul ales, de exemplu, pentru a urma trotuarele sau pentru a evita obstacolele, înainte de a se întoarce la acesta. Cea de-a patra etapă, care constă în recunoașterea destinației, completează procesul de navigare.

Un alt model de navigare umană, propune Adams¹² (1997), doar pentru pietoni. Acesta include 3 faze: planificarea preliminară a itinerariului, navigarea globală și navigarea fină.

Planificarea preliminară constă în a alege o destinație și a decide care este cel mai bun drum pentru a ajunge acolo. Ca și în modelul lui Downs and Stea¹³ (1977), această fază rezultă din mentalizarea unui itinerar de urmat, compus dintr-un anumit număr de puncte de decizie legate între ele prin segmente de drum mai mult sau mai puțin lungi. Regăsim aici aceiași factori de mediu (criterii legate de distanță, de timp al parcursului, de cunoaștere a priori a itinerariului) și subiectivi (noțiuni de securitate, de plăcere sau de stres, de exemplu) în ceea ce privește alegerea itinerariului. Întâlnirile cu formatori în domeniul locomoției și cu persoane cu deficiențe vizuale arată că aceste noțiuni subiective sunt deosebit de importante în cazul navigării persoanelor cu deficiențe vizuale.

Adams separă, apoi, navigarea în două faze numite navigare fină și globală.

Navigarea globală constă în a lega între ele diferitele puncte de reper care pot reprezenta dificultăți de locomoție (de exemplu o traversare a unei străzi sau o scară) sau intersecții, apoi a alege direcția care trebuie urmată în intersecții.

Navigarea fină este o sarcină imediată care constă în înfruntarea obstacolelor ivite în calea locomoției, indiferent dacă sunt prevăzute sau neprevăzute. Aceste

¹¹ Sabine Timpf, Gary Volta, David Pollock & Max Egenhofer, *A conceptual model of wayfinding using multiple levels of abstraction*, Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, ed. by A. Frank, I. Campari & U. Formentini, 348-67: Springer Berlin / Heidelberg, 1992.

¹² Adams, Christi J., *An investigation of navigation processes in human locomotion behavior*, Master of science Virginia: Polytechnic Institute and State University, 1997.

¹³ Downs, Roger M. & David Stea, *Maps in minds: reflections on cognitive mapping*, Harper & Row, 1977.

obstacole pot fi de ordinul unei trepte de trecut, dar și de un alt nivel, cum ar fi traversarea unei străzi care include, de exemplu, o trecere de pietoni, un semafor și traficul rutier. Navigarea fină este însoțită, în general, de o schimbare a strategiei locomotorii (în sensul biomecanic al termenului). Adams a remarcat, la subiecții văzători, că fazele de navigare fină necesită mai multă atenție decât fazele de navigare globală.

Aceste două modele nu se situează exact la același nivel și sunt complementare. Faza de planificare propusă de Adams¹⁴ (1997) conține fazele de orientare și de alegere a itinerariului ale lui Downs and Stea¹⁵ (1977).

Etapele următoare ale lui Downs and Stea (păstrarea itinerariului, recunoașterea destinației) sunt incluse, ambele, în navigarea globală a lui Adams. Noțiunea care apare în acest moment este cea de navigare fină, care capătă o importanță deosebită la persoanele cu deficiențe vizuale deoarece se referă la schimbarea strategiei locomotorii atunci când pietonul trebuie să depășească anumite dificultăți.

În aceste două modele, este important ca navigatorul să își poată actualiza poziția și orientarea pe măsură ce se deplasează. După cum am văzut, reperele vizuale sunt foarte utile. Totuși, există alte metode care permit integrarea acestor informații în timp (Loomis¹⁶ et al. 1998; Gallistel¹⁷ 1990).

Se pot cita aceste metode clasificate în funcție de cinematică. Navigarea bazată pe poziție (numită „pilotare”) are la bază semnale externe care indică poziția și orientarea călătorului. Acest tip de navigare se realizează, în general, cu ajutorul unei hărți externe sau interne. Navigarea bazată pe viteză (numită „estimare dedusă” engl. *dead/ded-reckoning* sau „integrarea căii urmate”, *path-integration*) este fundamentată pe semnale externe sau semnale proprioceptive (senzații interne legate de mișcare), care indică viteza călătorului. Deplasarea și direcția în raport cu punctul de origine sunt obținute prin integrarea vectorului viteză. În sfârșit, navigarea bazată pe accelerare (numită „navigare inerțială”) implică o dublă integrare a accelerărilor liniare și angulare ale călătorului, pentru a deduce de aici schimbările de poziție și de direcție survenite de la plecare. În acest caz, nu este necesar niciun semnal extern.

2.2.2 Cognația spațială a persoanelor cu deficiențe vizuale

Ca urmare a deficienței vizuale, procesele cogniției spațiale, în general, și ale navigării, în special, devin deosebit de complexe. Vederea furnizează informații spațiale multiple, simultane și într-o perioadă de timp foarte scurtă având în special, proprietatea de a stabiliza reprezentarea mentală a mediului în momentul deplasării (Thinus-Blanc & Florence Gaunet¹⁸ 1997).

¹⁴ Adams, Christi J., *An investigation of navigation processes in human locomotion behavior*, Master of science Virginia: Polytechnic Institute and State University, 1997.

¹⁵ Downs, Roger M. & David Stea, *Maps in minds: reflections on cognitive mapping*, Harper & Row, 1977.

¹⁶ Loomis, Jack M., Reginald G. Golledge & Roberta L. Klatzky, *Navigation System for the Blind: Auditory Display Modes and Guidance*, Presence-Teleoperators and Virtual Environments 7.193-203, 1998.

¹⁷ Gallistel, Charles R., *The organization of learning*, The MIT Press, 1990.

¹⁸ Thinus-Blanc, Catherine & Florence Gaunet, *Representation of space in blind persons: vision as a spatial sense?*, Psychol Bull 121.20-42, 1997.

2.2.2.1 Impactul lipsei vederii asupra cogniției spațiale

Fletcher (Fletcher¹⁹ 1980) a clasificat cercetările cu privire la capacitățile spațiale ale orbilor în trei mari teorii: deficiența (deficiency), ineficiența (inefficiency) și diferența (difference).

Teoria deficienței, născută din unele lucrări mai vechi ale lui von Senden, stipulează că orbii din naștere sunt incapabili să dezvolte o înțelegere generală a spațiului, deoarece nu au avut niciodată experiența proceselor vizuale necesare pentru a înțelege aranjamentele spațiale complexe. Această teorie are un interes istoric numai datorită faptului că numeroase lucrări mai recente au discreditat-o.

Teoria ineficienței afirmă că persoanele cu deficiențe vizuale pot înțelege și manipula mental concepte spațiale, însă, deoarece informația este bazată numai pe indici auditivi și haptici, această cunoaștere este inferioară celei bazate pe vedere (Rieser²⁰ et al. 1986: de exemplu). Orbii din naștere au tendința de a reconstrui spațiile parcurse ca itinerare liniare constituite dintr-o secvență de drumuri legate între ele prin puncte de decizie (Casey²¹ 1978). Această teorie sugerează că extrapolarea acestor drumuri la o reprezentare globală este dificilă, chiar imposibilă.

Teoria diferenței sugerează că persoanele cu deficiențe vizuale posedă aceleași capacități de a trata și de a înțelege concepte spațiale și că, deși echivalente din punct de vedere funcțional, acestea sunt elaborate diferit și, deseori, mai lent (Juurmaa²² 1973). Passini²³ et al (1988) afirmă că diferențele observate în ceea ce privește anumite sarcini spațiale între văzători și orbi pot fi explicate prin variabile indirecte, cum ar fi accesul la informație (de exemplu, hărți), experiența (văzătorii sunt mai experimentați în elaborarea de schițe sau de modele) sau stresul. Millar²⁴ (1994) susține că, deși simțurile nevizuale sunt „inferioare” pentru a coda informația spațială, persoanele cu deficiențe vizuale au același potențial ca și văzătorii pentru a dezvolta o reprezentare integrată a spațiului. Această afirmație este confirmată de numeroase lucrări care arată că persoanele cu deficiențe vizuale sunt capabile să construiască hărți cognitive (a se vedea, de exemplu Casey²⁵ 1978; Dodds²⁶ et al. 1982; Hollyfield & Foulke²⁷ 1983).

¹⁹ Fletcher, Janet F., *Spatial Representation in Blind Children. 1: Development Compared to Sighted Children*, Journal of Visual Impairment and Blindness 74.381-85, 1980.

²⁰ Rieser, John J., David A. Guth & Everett W. Hill, *Sensitivity to perspective structure while walking without vision*, Perception 15.173-88, 1986.

²¹ Casey, Steven M., *Cognitive mapping by the blind*, Journal of Visual Impairment & Blindness 72.297-301, 1978.

²² Juurmaa, Jyrki, *Transportation in Mental Spatial Manipulation: A Theoretical Analysis*, American Foundation for the Blind Research Bulletin 26.87-143, 1973.

²³ Passini, Romedi & Guylène Proulx, *Wayfinding without vision*, Environment and Behavior 20.227-52, 1988.

²⁴ Millar, Susanna, *Understanding and representing space: Theory and evidence from studies with blind and sighted children*, Oxford University Press, USA, 1994.

²⁵ Casey, Steven M., *Cognitive mapping by the blind*, Journal of Visual Impairment & Blindness 72.297-301, 1978.

²⁶ Dodds, Allan G., Ian Howarth & David Clark-Carter, *The mental maps of the blind: The role of previous visual experience*, Journal of Visual Impairment & Blindness 76.5-12, 1982.

²⁷ Hollyfield, Rebecca L. & Emerson Foulke, *The spatial cognition of blind pedestrians*, Journal of Visual Impairment & Blindness, 1983.

2.2.2.2 Influența vârstei de apariție sau a duratei cecității

În prezent, discuția este orientată mai degrabă către teoria diferenței, în care vârsta apariției sau durata cecității ar putea avea o influență asupra cunoștințelor spațiale și a mecanismelor care permit dobândirea acestora. Nevăzătorii tardivi (NVT) au avut o experiență vizuală care ar trebui să le permită să beneficieze de un avantaj în ceea ce privește cogniția spațială față de nevăzătorii precoci sau congenitali (NVP). Într-adevăr, s-a observat că subiecții văzători cu ochii acoperiți și NVT reușesc mai bine decât NVP în majoritatea sarcinilor de navigare.

Se pare că există diferențe cognitive între orbii congenitali, orbii tardivi și văzători, dar că, în general, acestea nu au consecințe asupra realizării sarcinilor spațiale. Diferențele observate sunt legate, probabil, de o dificultate de învățare a tehnicilor de explorare a spațiului și un acces mai mic la informațiile externe în cazul persoanelor cu deficiențe vizuale. Probabil că lipsa consecințelor asupra realizării de sarcini spațiale rezultă din folosirea intensivă a altor modalități senzoriale și din punerea în aplicare a unor moduri diferite de explorare.

2.2.2.3. Dificultăți, necesități și factori care influențează navigarea persoanelor cu deficiențe vizuale

Odată cu progresul tehnologiilor de asistare pentru comunicarea scrisă, unul dintre cele mai mari impacturi ale deficienței vizuale este, fără nicio îndoială, pierderea mobilității autonome. Potrivit unui studiu efectuat de Ministerul Sănătății din Franța (Sander²⁸ et al. 2005), 58% dintre persoanele cu deficiențe vizuale care au o condiție fizică ce le permite să se deplaseze au declarat că întâmpină probleme în cazul deplasărilor în exterior. 29% au declarat că nu sunt capabili să se deplaseze singuri, iar 15% pot să se deplaseze singuri, dar numai pe itinerare cunoscute. De asemenea, deficiența vizuală este o frână în ceea ce privește mobilitatea în interior pentru 40% dintre subiecți.

Pentru persoanele cu deficiențe vizuale, navigarea reprezintă o sarcină relativ complexă de realizat, care generează un stres important. Într-adevăr, vederea furnizează nu numai informații cu privire la deplasările efectuate, ci și în ceea ce privește dispunerea spațiului apropiat și îndepărtat (Foulke²⁹ 1982; Foulke³⁰ 1971 ; Strelow³¹ 1985). În consecință, în ciuda utilizării câinelui ghid sau a bastonului alb, navigarea rămâne aproape imposibilă și periculoasă atunci când traseul este necunoscut sau nu cuprinde suficiente indicii nevizuale (de exemplu, în zonele nou delimitate).

²⁸ Sander, Marie-Sylvie, Marie-Christine Bournot, Françoise Lelièvre & Anne Tallec, *Les personnes ayant un handicap visuel. Apports de l'enquête handicaps - incapacités - dépendances* (ed.) R.d.e.d.M.F.d.I.S.e.d. Solidarités, 2005.

²⁹ Foulke, Emerson, *Perception, cognition, and the mobility of blind pedestrians*. Spatial abilities: Development and physiological foundations, ed. by M. Potegal, 55-76; New York: Academic Press, 1982.

³⁰ Foulke, Emerson, *The perceptual basis for mobility*, American Foundation for the Blind, Research Bulletin, 1971.

³¹ Strelow, Edward W., *What is needed for a theory of mobility: Direct perceptions and cognitive maps—lessons from the blind*, Psychological review 92.226, 1985.

Dificultățile variază în funcție de mediu, de densitatea populației, de climă, de constrângerile arhitecturale, de absența reperelor etc., și sunt percepute diferit în funcție de individ. Deși toate modalitățile senzoriale (auz, pipăit, miros, „masă”³²) sunt mobilizate în timpul unei deplasări, acestea nu sunt suficiente întotdeauna la persoanele cu deficiențe vizuale pentru a se deplasa în mod sigur și autonom (Wolff³³ et al. 2006). Dejeammes³⁴ și colaboratorii (2008) au identificat patru nevoi primordiale legate de fiecare deplasare:

- (1) securitatea (evitarea accidentelor de circulație, evitarea căderilor și a șocurilor);
- (2) localizările personale (unde sunt, pe ce, între ce și ce ...?) și relative (strada / străzile / clădirea / magazinul / stația de transport X este oare la dreapta / la stânga / în față?);
- (3) orientarea (oare merg în direcția bună?);
- (4) informația (ce este în jurul meu în ceea ce privește infrastructura / activitățile / informațiile scrise disponibile?).

Aceste diferite nevoi vor fi satisfăcute datorită utilizării unor indicii nevizuale cu privire la parcurs, puse în legătură cu hărțile mentale preexistente. Strategii suplimentare cum ar fi numărarea în cazul distanței (de exemplu, numărul de pași efectuați), luarea în considerare a timpului (durata scursă) și a etapelor de parcurs (de exemplu, a o lua la dreapta după cea de-a treia stradă) vor permite efectuarea deplasărilor diminuând stresul.

2.2.2.4 Modurile de compensare senzorială

În general, se admite că persoanele oarbe compensează lipsa informației vizuale printr-o utilizare mai mare a informațiilor furnizate de sistemele lor senzoriale intacte. Cu toate acestea, datele empirice rămân foarte contradictorii. Această incoerență se poate datora, în parte, diversității participanților din diferitele studii (importanță, durată, etiologia deficienței; vârsta subiecților, tipul sau lipsa unui grup martor corespunzător), utilizării unor sarcini ce nu sunt complet identice etc. (a se vedea Millar³⁵ 1982; Roder & Neville³⁶ 2003; Thinus-Blanc & Florence Gaunet³⁷ 1997; pentru o discuție detaliată).

³² Simțul maselor nu este un al șaselea simț fiziologic. Este o abilitate dobândită, de a resimți prezența unei mase mai mult sau mai puțin importante (zid, coloană, streășină) sau discontinuitatea acesteia (de exemplu, prezența unui culoar). Aceasta se datorează, probabil, percepției simultane a unor indicii auditive și somestezice foarte slabe.

³³ Wolff, Marion, Philippe Cabon, Gérard Uzan, Julien Nelson & Stanislas Couix, *Déplacement urbain de personnes non-voyantes: étude multi-factorielle des difficultés et apport d'une nouvelle interface pour le recueil des données*, Paper presented at the Ergo-IA 2006, Biarritz, 2006.

³⁴ Dejeammes, Maryvonne, Gérard Uzan, M'Balo Seck & Catherine Sidot, *Déplacements des déficients visuels en milieu urbain: analyse des besoins en sécurité, localisation et orientation et pistes d'évolution*, CERTU, 2008.

³⁵ Millar, Susanna, *Studies of the deaf and the blind*, The Pathology and psychology of cognition, 135: Taylor & Francis, 1982.

³⁶ Roder, Brigitte & Helen Neville, *Developmental functional plasticity. Handbook of neuropsychology*, ed. by J. Grafman & I.H. Robertson, 231-70: Elsevier, 2003.

³⁷ Thinus-Blanc, Catherine & Florence Gaunet, *Representation of space in blind persons: vision as a spatial sense?*, Psychol Bull 121.20-42, 1997.

S-ar părea că facultățile perceptive periferice nu sunt mai bune la persoanele cu deficiențe vizuale. Cu toate acestea, unele procese – cum ar fi învățarea perceptivă, factorii atenționali, interacțiunile multimodale, plasticitatea cerebrală – ar putea să le permită să depășească văzătorii într-un anumit număr de sarcini haptice sau auditive. Din contră, aceștia ar fi mai slabi în unele sarcini în care vederea joacă un rol important (a se vedea Zwiers³⁸ et al. 2001a; Zwiers³⁹ et al. 2003).

2.2.2.5 Navigarea fără vedere

Pentru a studia navigarea în absența vederii, Loomis⁴⁰ a definit cinci procese generale: (1) detectarea, (2) memorarea unui traseu, (3) stabilirea unei reprezentări globale a mediului, (4) selectarea itinerarelor și (5) realizarea acestor itinerare.

Primul proces (detectarea) se referă la dobândirea unor informații cu privire la propria deplasare și/sau la repere înconjurătoare, oricare ar fi modalitatea senzorială implicată.

Memorarea unui traseu corespunde unei secvențe de segmente, reorientări, depinde de denaturări de percepție și de eventuale pierderi de memorie.

Cel de-al treilea proces are la bază informațiile dobândite în timpul deplasării pentru a elabora o reprezentare spațială globală de tipul hărții. Informațiile incluse aici ar putea fi obiecte întâlnite, puncte de interes sau deosebit de importante.

Cel de-al patrulea proces permite alegerea unor itinerare pe baza memorării unui traseu sau a unei reprezentări globale.

Cel de-al cincilea procedeu permite efectuarea deplasării.

Fiecare dintre componentele navigării pedestre va deveni dificilă prin deficiența vizuală. Se observă o diminuare a sentimentului de securitate, în raport cu o lipsă aproape totală a anticipării perceptive, precum și o dificultate importantă de a păstra o orientare corectă, ca urmare a diminuării reperelor. Lipsa vederii reduce percepția informațiilor cu privire la mediu, ceea ce are drept consecință o dihotomie între câmpul apropiat și cel îndepărtat. Astfel, dobândirea de informații cu privire la câmpul apropiat are la bază, în principal, modalitatea somestezică care include informații tactile, termice, proprioceptive (percepția de sine), chinestezice (percepția propriei deplasări) și nociceptive (percepția durerii). Sursa cea mai bogată rămâne percepția tactilă care poate fi pasivă sau activă (numită și percepție haptică). Informațiile calitative olfactive sau auditive pot fi completate și de informațiile somestezice. Accesul la câmpul îndepărtat este posibil numai prin auz. Auzul nefiind o modalitate completă la nivel spațial (nici o informație de configurare, de exemplu), dobândirea de informații privind câmpul îndepărtat este completată de locomotie. Acest mod de percepție obligă la o explorare secvențială a mediului, mai lentă și mai costisitoare. Informațiile spațiale vor fi integrate mai mult egocentric decât alocentric

³⁸ Zwiers, Marcel P., John A. Van Opstal & Johan R. Cruysberg, *A spatial hearing deficit in early-blind humans*, Journal of Neuroscience 21.RC142-RC45, 2001a.

³⁹ Zwiers, Marcel P., John A. Van Opstal & Gary D. Paige, *Plasticity in human sound localization induced by compressed spatial vision*, Nature Neuroscience 6.175-81, 2003.

⁴⁰ Loomis, Jack M., Roberta L. Klatzky, Reginald G. Golledge, Joseph G. Cicinelli, James W. Pellegrino & Phyllis A. Fry, *Nonvisual navigation by blind and sighted: assessment of path integration ability*, J.Exp.Psychol.Gen. 122.73-91, 1993.

(Millar⁴¹ 1994). Persoanele cu deficiențe vizuale elaborează, reprezentări mentale de tip drum. Experiențele lui Tinti⁴² et al (2006) temperează aceste afirmații, arătând că experiența vizuală nu este absolut necesară pentru dezvoltarea reprezentărilor spațiale complexe de tip hartă.

2.2.3 De ce se studiază cogniția spațială a persoanelor cu deficiențe vizuale?

Deși vederea este simțul cel mai eficient pentru a obține informații spațiale, persoanele oarbe sunt capabile să utilizeze indicii senzoriale (auditive, olfactive sau somestezice) sau informații scrise, desenate sau orale pentru a dobândi o cunoaștere a mediului lor geografic.

Studierea cogniției spațiale a persoanelor cu deficiențe vizuale are mai multe avantaje, teoretice și practice. Patru dintre acestea par mai importante. Mai întâi, pentru a învăța un mediu geografic pornind de la surse primare (cele prezente în mediul înconjurător) sau secundare (prin intermediul hărților, de exemplu), vederea este simțul cel mai eficient. Posibilitatea de a se deplasa sau de a interacționa cu mediul într-un mod autonom este, deci, una dintre marile provocări cu care se confruntă persoanele cu deficiențe vizuale (Golledge⁴³ 1993). Bruce et al. (Bruce⁴⁴ et al. 1991) arată, de exemplu, că în săptămâna anterioară unei anchete realizate pe un eșantion de tineri orbi, 20 la sută dintre respondenți nu își părăsiseră domiciliul, 34 la sută se deplasaseră pe un perimetru foarte restrâns și numai 41 la sută părăsiseră incinta casei singuri și pe jos.

Clark-Carter et al. (Clark-Carter⁴⁵ et al. 1986) au raportat, de asemenea, că cel puțin 30 la sută dintre persoanele cu deficiențe vizuale nu fac nicio deplasare singure în afara casei. Atunci când se deplasează, persoanele cu deficiențe vizuale folosesc în principal itinerare familiare și cunoscute.

Într-adevăr, explorarea poate fi o sursă a dezorientării și a anxietății. Cercetările în domeniul cartografiei cognitive au potențialul de a ameliora calitatea vieții persoanelor cu deficiențe vizuale furnizând indicii pentru a ameliora formările în ceea ce privește orientarea și mobilitatea și de a transforma, astfel, subiecții cu deficiențe vizuale, în persoane mai autonome.

În al doilea rând, cunoștințele obținute datorită cercetărilor în domeniul cartografiei cognitive sunt utile pentru a concepe mediile înconjurătoare (în special urbane) mai ușor de memorat și mai plăcut de parcurs. Modificări ar putea fi aduse,

⁴¹ Millar, Susanna, *Understanding and representing space: Theory and evidence from studies with blind and sighted children*, Oxford University Press, USA, 1994.

⁴² Tinti, Carla, Mauro Adenzato, Marco Tamietto & Cesare Cornoldi, *Visual experience is not necessary for efficient survey spatial cognition: evidence from blindness*, *Q.J.Exp.Psychol* (Colchester.) 59.1306-28, 2006.

⁴³ Golledge, Reginald G., *Geography and the disabled: a survey with special reference to vision impaired and blind populations*, *Transactions of the Institute of British Geographers* 18.63-85, 1993.

⁴⁴ Bruce, Ian, Aubrey McKennel & Errol Walker, *Blind and partially sighted adults in Britain: The RNIB Needs Survey*, Volume 1, HMSO, London, 1991.

⁴⁵ Clark-Carter, David D., Anthony D. Heyes & Ian Howarth, *The efficiency and walking speed of visually impaired people*, *Ergonomics* 29, 1986.

în special, mediilor existente pentru a permite un acces mai bun persoanelor cu deficiențe vizuale.

În al treilea rând, conceperea unor dispozitive de asistare pentru persoanele cu deficiențe vizuale a progresat foarte mult în ultimii ani. În afara dispozitivelor de ajutor pentru mobilitate care permit evitarea obstacolelor imediate, în ultimul timp s-au dezvoltat numeroase dispozitive de ajutor pentru orientare, care au la bază poziționarea geografică precisă și ghidarea persoanelor cu deficiențe vizuale (Farcy⁴⁶ et al. 2006). Pentru ca aceste dispozitive să fie utilizabile, este esențială integrarea persoanelor cu deficiențe vizuale în verigile de concepție pentru a înțelege practicile „clasice” de navigare ale acestora, dar și nevoile lor și modul în care ei utilizează aceste sisteme (Anke Brock⁴⁷ 2010). Cercetările în domeniul cogniției spațiale pot permite să se înțeleagă ce informație este necesară pentru un navigator orb și cum trebuie să i se prezinte acestuia informația. Cercetările în domeniul cartografiei cognitive pot fi, de asemenea, foarte utile pentru a dezvolta metode de evaluare ale acestor dispozitive de asistare.

În al patrulea rând, studierea cogniției spațiale a persoanelor cu deficiențe vizuale aduce informații prețioase cu privire la rolul experienței senzoriale în cartografia cognitivă în general. Acest lucru conduce la progrese teoretice care pot avea repercusiuni în alte domenii în care cogniția spațială este importantă, cum ar fi robotica.

Lucrările descrise aici privesc numai ultimele două puncte (conceperea unor dispozitive de supleanță și aspectele teoretice care privesc cogniția spațială, comparată în special cu aceea a văzătorilor) și sunt realizate în strânsă colaborare cu Centrul de Educație Specializată pentru Persoanele cu Deficiențe Vizuale – Institutul Tinerilor Orbi din Toulouse (Centre d'Éducation Spécialisé pour Déficients Visuels – Institut des Jeunes Aveugles de Toulouse).

2.2.4 Asistarea navigării și terminologie

Lucrările descrise cu privire la conceperea unor sisteme de asistare a navigării se sprijină pe etapele navigării descrise de Downs and Stea⁴⁸ (1973). Atunci când este necesar, se utilizează și noțiunile de planificare, navigare fină și globală, introduse de Adams⁴⁹ (1997) care accentuează dificultățile generate de sarcinile de locomoție care nu sunt ghidate vizual. În acord cu terminologia utilizată de specialiștii din domeniul readaptării nevăzătorilor (deși nu total echivalenți), se vor utiliza mai frecvent termenii de „mobilitate”, „orientare” și „pregătire a itinerariului”.

⁴⁶ Farcy, René, Roger Leroux, Alain Jucha, Roland M. Damaschini, Colette Gregoire & Aziz Zogaghi, *Electronic travel aids and electronic orientation aids for blind people: technical, rehabilitation and everyday life points of view*, CVHI, 2006.

⁴⁷ Brock, Anke, *Cartes interactives multimodales pour piétons malvoyants*, Master thesis Univ. P. Sabatier, Toulouse, 2010.

⁴⁸ Downs, Roger M. & David Stea, *Image & environment: Cognitive mapping and spatial behavior*, Edward Arnold, London, 1973.

⁴⁹ Adams, Christi J., *An investigation of navigation processes in human locomotion behavior*, Master of science Virginia: Polytechnic Institute and State University, 1997.

Mobilitatea corespunde facultății de a se deplasa păstrând o direcție și evitând obstacolele. Aceste două abilități de a păstra o direcție și de a evita obstacolele sunt predate încă de la începutul cursurilor de formare în domeniul locomotiei și permit persoanelor cu deficiențe vizuale să se concentreze pe probleme de orientare.

Orientarea corespunde facultății de a se situa în mediul înconjurător, precum și de a-și situa punctul de plecare, punctul de sosire și toate punctele intermediare necesare pentru a ajunge la destinație. Conform modelor descrise anterior, această fază de orientare în navigare a persoanelor cu deficiențe vizuale se bazează foarte mult pe pregătirea itinerariului, prezența unor puncte de reper și alegerea direcțiilor.

Pregătirea unor itinerare pare a fi primordială, deoarece majoritatea persoanelor cu deficiențe vizuale resimt stres și anxietate la ideea de a se deplasa într-un mediu necunoscut. Pregătirea itinerariului include etapele preliminare de dobândire a unor cunoștințe spațiale prin intermediul hărților, de exemplu, precum și faza de planificare care permite alegerea itinerariului în funcție de factorii subiectivi și externi. Prezența unor puncte de reper nevizuale în itinerariul ales are o importanță capitală. Aceste puncte de reper vor servi, în același timp, ca puncte de decizie (pentru o schimbare de direcție, de exemplu), dar și ca puncte de confirmare care dau asigurări că s-a făcut alegerea bună. Punctele de interes (prezența unui parc, de exemplu) influențează în mod evident alegerea itinerariului, dar pot servi și ca puncte de reper. În cele din urmă, alegerea direcțiilor este, de asemenea, o etapă deosebit de complicată la persoanele cu deficiențe vizuale, deoarece trebuie să se traducă într-o modalitate nevizuală niște indicații care sunt, foarte frecvent, vizuale („a o lua pe strada V. Hugo” de exemplu). Această observație întărește avantajul punctelor de confirmare care trebuie să apară rapid după alegerile efectuate.

2.2.5 Tehnologiile de asistare pentru persoanele cu deficiențe vizuale

Ajutoarele tehnice pentru persoanele cu deficiențe vizuale au obiectivul de a compensa limitările de activitate sau restricțiile de participare legate de deficiența vizuală. Limitările de activitate desemnează dificultățile pe care o persoană le întâmpină, în executarea unor activități. Restricțiile de participare desemnează problemele pe care o persoană le poate întâlni în implicarea sa într-o situație de viață reală. În prezent, există numeroase dispozitive (care nu vor fi descrise aici), legate de accesibilitatea informației (lectură și scriere). În domeniul spațial, navigarea constituie miza cea mai importantă. Unele ajutoare tehnice, precum bastonul alb sau câinele ghid, au devenit standarde care au constituit mari progrese pentru autonomia persoanelor oarbe. Mai recent, s-au dezvoltat și numeroase ajutoare electronice, care au evoluat odată cu progresele legate de miniaturizare, de puterea de calcul, de informatica integrată, pervazivă etc.

În acest domeniu, trebuie să se distingă două tipuri de dispozitive:

- sistemele „generice” caută să substituie vederea în totalitatea sa (a se vedea White⁵⁰ et al. 1970; Bach-y-Rita⁵¹ et al. 1969b; Arno⁵² et al. 2001; Meijer⁵³ 1992);

⁵⁰ White, Benjamin, Frank A. Saunders, Lawrence Scadden, Paul Bach-Y-Rita & Carter C. Collins, *Seeing with the skin*, Attention, Perception, & Psychophysics 7.23-27, 1970.

- dispozitive specifice îndeplinirii unei sarcini, cum ar fi ajutoarele pentru navigare (a se vedea, de exemplu, Loomis⁵⁴ et al. 1994; Ran⁵⁵ et al. 2004) sau ajutoarele pentru explorarea unor locuri geografice prin intermediul unor hărți sau modele (Wright⁵⁶ et al. 2010).

Această disociere se înscrie într-o dezbatere cu privire la substituția senzorială, concept inventat în anii 1960 pentru a defini acțiunea de a substitui o modalitate senzorială printr-o alta, cum este neuroproteza vizuală. Acestea constituie tehnologii care au ca scop refacerea vederii conectând un captor artificial (pe principiul camerei video), la sistemul nervos.

2.2.5.1 Sistemele de vedere artificială

Sistemele electronice complementare, suplente, fie că sunt generice sau specifice, au o arhitectură asemănătoare. Într-adevăr, ele sunt concepute după același principiu: un lanț de achiziție a informației, un lanț de transformare a informației și un modul de restituire a informației analizate. În cazul sistemelor de vedere artificială, achiziția se face, în general, printr-o cameră integrată, montată pe o cască sau pe ochelari. Transformarea informației este foarte simplă deoarece se limitează, de exemplu, la corespondența, punct cu punct, a pixelilor imaginii către interfața de ieșire (cu o diminuare sistematică a numărului de puncte pentru a corespunde cu interfața de ieșire) sau la instalarea unui filtru spațial bazat pe contururi. Exemplele de sisteme de vedere artificială sunt numeroase în literatură și au ca modalitate de ieșire somestezia sau auzul. Aceste sisteme se bazează pe ipoteza că este posibil să se redea un număr mare de informații spațiale (forme și poziții) privind mediul înconjurător prin intermediul modalității de ieșire.

3. Conceperea unor tehnologii de supleanță, metode, instrumente

O nouă temă dezvoltată la IRIT, având ca obiect înțelegerea și modelarea percepției spațiale și cogniția în deficiențele de vedere, are scopul de a proiecta și de

⁵¹ Bach-y-Rita, Paul, Carter C. Collins, Benjamin White, Frank A. Saunders, Lawrence Scadden & Robert Blomberg, *A tactile vision substitution system*, Am.J.Optom.Arch.Am.Acad.Optom. 46.109-11, 1969.

⁵² Arno Patricia, Annick Vanlierde, Emmanuel Streel, Marie-Chantal Wanet-Defalque, Sandra Sanabria-Bohorquez & Claude Veraart, *Auditory substitution of vision: pattern recognition by the blind*, Applied Cognitive Psychology 15.509-19, 2001.

⁵³ Peter B. L. Meijer, *An experimental system for auditory image representations*, IEEE Trans.Biomed.Eng 39.112-21, 1992.

⁵⁴ Loomis, Jack M., Reginald G. Golledge, Roberta L. Klatzky, Jon M. Speigle & Jerome Tietz, *Personal guidance system for the visually impaired*, Marina Del Rey, California, United States, 1994.

⁵⁵ Ran, Lisa, Sumi Helal & Steve Moore, *Drishti: an integrated indoor/outdoor blind navigation system and service*, Paper presented to the Second IEEE International Conference on Pervasive Computing and Communications (PerCom'04), 2004.

⁵⁶ Wright, Tessa, Beth Harris & Eric Sticken, *A Best-Evidence Synthesis of Research on Orientation and Mobility Involving Tactile Maps and Models*, JVIB 104.95-106, 2010.

a evalua sistemele de asistare automată a cogniției care să permită creșterea autonomiei nevăzătorilor în sarcini spațiale. Acest proiect are un dublu obiectiv: furnizarea de sisteme, modele și instrumente pentru proiectarea HMI (Human Machine Interface) centrat pe utilizator, dar și studiul principiilor prin care un sistem de vedere artificială poate fi conectat direct la cortex (printr-o interfață neurală).

În încercarea de reconstrucție a unei scene vizuale, se abordează pentru început o sub-problemă și anume, simularea unor funcții utile în sistemul vizual pentru navigare: recunoașterea și localizarea obiectelor. Această caracteristică ne permite restaurarea comportamentelor video-motorii importante pentru asistența navigării asistate de calculator prin Internet. În conformitate cu analiza nevoilor persoanelor cu deficiențe de vedere, propunem dezvoltarea unor instrumente care să ajute la pregătirea traseului, off-line.

Se descriu în continuare instrumentele și metodele implementate la IRIT, mai multe prototipuri, inclusiv prototipul asistenței spațiale proiectate de echipa ELIPSE.

În mod paradoxal, utilizarea celor mai cunoscute dispozitive de asistare pentru mobilitate, bastonul alb și câinele ghid, nu este atât de curentă. În Franța, acestea nu au fost adoptate decât de 2% dintre persoanele cu deficiențe vizuale, 26% din populația oarbă (Sander⁵⁷ et al. 2005). Utilizarea limitată a câinilor ghid se poate explica prin costul său ridicat: aproximativ 15000€ în total, incluzând și o perioadă de muncă de 8 ani⁵⁸. Bastonul alb nu este atât de costisitor (aproximativ 50€) și oferă informații cu privire la mediul din imediata apropiere a utilizatorului. Pietonii cu deficiențe vizuale care îl utilizează sunt ușor de identificat, deci vor fi mai protejați. Principala limitare a bastonului alb în cadrul mobilității este mărimea acestuia: distanța de aproximativ doi pași pe care o acoperă nu permite anticiparea drumului de urmat și limitează viteza de mers. În plus, bastonul alb nu împiedică coliziunile cu obstacolele situate la înălțime. În sfârșit, deși poate fi utilizat pentru a verifica prezența reperelor, acesta nu aduce niciun ajutor în orientare. Pentru a depăși aceste limitări, au fost dezvoltate numeroase aparate electronice dedicate mobilității și orientării, însă acestea nu sunt foarte răspândite (Dowling⁵⁹ 2003).

Aceste dispozitive au fost concepute în mod frecvent în baza unei tehnologii, urmând a fi adaptate ulterior pentru a răspunde nevoilor persoanelor cu deficiențe vizuale. Acest demers de concepție a dus la crearea unor dispozitive de supleanță care funcționează, dar care nu sunt utilizate deoarece sunt concepute prost din punctul de vedere al facilității de utilizare. Echipa ELIPSE a IRIT are o experiență îndelungată în ceea ce privește concepția participativă, mai ales în cadrul concepției sistemelor de supleanță. În colaborare cu Institutul Tinerilor Orbi din Toulouse, echipa a utilizat o metodă de concepție modulară bazată pe agenți autonomi, comunicând prin schimb de mesaje pe o magistrală logică.

⁵⁷ Sander, Marie-Sylvie, Marie-Christine Bournot, Françoise Lelièvre & Anne Tallec, *Les personnes ayant un handicap visuel. Apports de l'enquête handicaps - incapacités - dépendances* (ed.) R.d.e.d.M.F.d.I.S.e.d, Solidarités, 2005.

⁵⁸ Sursa : Asociația Cânilor Ghizi pentru Orbi, <http://www.chien-guide.org> (2008)

⁵⁹ Dowling, Jason, *Mobility enhancement using simulated artificial human vision*, PhD: Queensland University of Technology, 2003.

3.1 Colaborarea cu persoanele cu deficiențe vizuale și cu profesioniștii în recuperare

Unul dintre defectele de care suferă, în prezent, metodele de concepție a sistemelor de supleanță este lipsa de implicare a utilizatorilor finali. Încă de la primele faze de modelizare a aplicației finale, concepția trebuie gândită integrând etapele de evaluare și de revenire la concepție. Pentru a optimiza acest demers, Christophe Jouffrais a creat, timp de 5 ani, legături solide cu Centrul de educație specializată pentru persoane cu deficiențe vizuale IJA – Institut des Jeunes Aveugles de Toulouse / Institutul Tinerilor Orbi din Toulouse⁶⁰.

Cu sprijinul consiliului de administrație al IJA, s-a elaborat o convenție pentru fondarea unui laborator de cercetare comun, care va permite persoanelor cu deficiențe vizuale, profesioniștilor în domeniul formării și recuperării și cercetătorilor IRIT să se reunească în jurul unor discuții și proiecte comune. De asemenea, pentru colectarea de fonduri destinate cercetării în domeniul tehnologiei de supleanță pentru persoanele cu deficiențe vizuale, s-a realizat un proiect de întemeiere a unei fundații.

În diferitele proiecte de cercetare prezentate în continuare, echipa ELIPSE apelează în mod frecvent la IJA pentru a defini nevoile, dar și în scopul de a recruta un grup de subiecți pentru experiențele de psihologie experimentală și/sau pentru fazele de concepție⁶¹.

3.2 Concepția participativă cu persoanele cu deficiențe vizuale

Concepția participativă este un proces de concepție a sistemelor interactive care implică utilizatorii în întregul proces al dezvoltării (Norman & Draper⁶² 1986). Totuși, aceasta presupune ca utilizatorii să dispună de toate capacitățile lor fizice, în special vizuale. În cercetările sale, echipa ELIPSE a remarcat că metodele și instrumentele utilizate în mod tradițional nu sunt adaptate pentru persoane deficiente vizual și, în consecință, a făcut o serie de observații și recomandări care permit adaptarea metodelor de concepție participativă la utilizatorii cu deficiențe vizuale (Anke Brock⁶³ et al. 2010b).

3.3 Concepție modulară și prototipare rapidă

Dezvoltarea de prototipuri funcționale ale unor sisteme interactive complexe este o adevărată problemă în informatică, mai ales atunci când aceste dispozitive

⁶⁰ CESDV-IJA este un centru de educație specializat pentru deficienți vizuali. Fondat în 1866, instituția primește copii orbi și cu probleme de vedere de ambele sexe, în regim de demi-pensiune, internat sau externi, și adulți care participă la o formare profesională. A se vedea <http://www.ija-toulouse.cict.fr/>

⁶¹ Se face aici o diferență între experiențe și evaluări. Experiențele sunt realizate înaintea oricărei concepții în scopul de a observa comportamentele, capacitățile perceptive sau cognitive ale deficienților vizuali. Evaluările privesc testele făcute pe prototipul unui dispozitiv de supleanță.

⁶² Norman, Donald A. & Stephen W. Draper, *User centered system design* Hillsdale, NJ, 1986.

⁶³ Brock, Anke, Jean-Luc Vinot, Bernard Oriola, Slim Kammoun, Philippe Truillet & Christophe Jouffrais, *Méthodes et outils de conception participative avec des utilisateurs non-voyants*, Paper presented to the IHM'10, 2010b.

se adresează unor subiecți lipsiți de vedere (ca urmare a absenței interfeței grafice, care permite compensarea a numeroase probleme tehnice). Adesea, tehnica este o frână pentru concepția participativă, cu feed-back-uri prea lente în ceea ce privește integrarea ultimelor rezultate și a noilor idei. Majoritatea mediilor de dezvoltare sunt foarte limitate la platforme software sau la materiale specifice. Pentru a atenua aceste dificultăți, se disting două tendințe: alegerea unui singur mediu de dezvoltare coerent, însă cu prețul unor potențialități restrânse sau operarea pe diferite platforme, dar cu problemele de comunicare și de integrare ale acestor dezvoltări în cadrul unui singur sistem interactiv performant.

Echipa ELIPSE, în colaborare cu CENA (centrul de studii pentru navigare aeriană), a dezvoltat o metodă de prototipare rapidă bazată pe o magistrală logică (Buisson⁶⁴ et al. 2002). Această magistrală, denumită Ivy⁶⁵, funcționează prin schimb de mesaje textuale de la un agent la altul. Nicio structură de date complexă nu poate fi trimisă. Această restricție numai la lanțurile de caractere face ca datele să fie compatibile sub forma unui standard disponibil pe toate platformele și în toate limbajele. Agenții trimit și primesc mesaje pe o adresă de broadcast. Astfel, fiecare agent se poate abona și poate asculta mesaje filtrate prin intermediul unui prefix pentru mesaje și poate invoca o funcție evenimentială la fiecare receptare de mesaje. Datorită acestei funcționări, este posibil să se suprimă un modul, să fie simulat sau înlocuit foarte ușor. Singurul imperativ este ca modulele să fie pe aceeași sub-rețea.

Această metodă permite adăugarea și/sau suprimarea agenților în dispozitive, fără a le denatura funcționarea globală. În toate proiectele, echipa ELIPSE dezvoltă module independente (de exemplu, pentru achiziționarea unor date video sau pentru redarea sonoră) și autonome, care pot fi testate și simulate separat. Această funcționare facilitează evaluarea independentă a fiecărui modul și permite evaluarea pe aceeași bază software a unor metode diferite de achiziție, de transformare și de redare, fără a schimba arhitectura sistemului. Mai mult, modulele sunt reutilizabile în diferite dispozitive. Această metodă permite, în special, concentrarea atenției asupra concepției centrate pe utilizator (Norman & Draper⁶⁶ 1986) suprimând unele bariere tehnice.

3.4 Platforme, metode și instrumente

La IRIT, există două platforme dedicate studiului realității virtuale: PREVI, (<http://www.irit.fr/-PREVI->) și ULYSS (<http://www.irit.fr/labo-usages/>). Aceste două platforme permit utilizarea ultimelor tehnologii interactive și imersive, precum și a metodelor și instrumentelor de concepere a unor sisteme interactive. S-a inițiat înstituirea unui post experimental (numit Human Experimentation Lab) în care se pot dezvolta noi protocoale privind studiul percepției auditive sau haptice și studiul simulării unei neuroproteze vizuale. Acest post este prevăzut cu o cască pentru

⁶⁴ Buisson, Marcellin, Alexandre Bustico, Stéphane Chatty, François-Régis Colin, Yannick Jestin, Sébastien Maury, Christophe Mertz & Philippe Truillet, *Ivy: un bus logiciel au service du développement de prototypes de systèmes interactifs*, 2002.

⁶⁵ Ivy este dezvoltat sub licența LGPL de către DTI/SDER. A se vedea <http://www.tls.cena.fr>

⁶⁶ Norman, Donald A. & Stephen W. Draper, *User centered system design* Hillsdale, NJ, 1986.

realitatea virtuală dotată cu un oculometru (NVisor + SR research), ceea ce permite ajustarea afișajului în funcție de poziția ochilor, cu un sistem de captură a mișcării (Optitrack de NaturalPoint, 12 camere la 100 Hz) și cu o platformă ce conține 35 HP (Figura 2), care permite studierea localizării spațiale auditive în câmpul peripersonal.

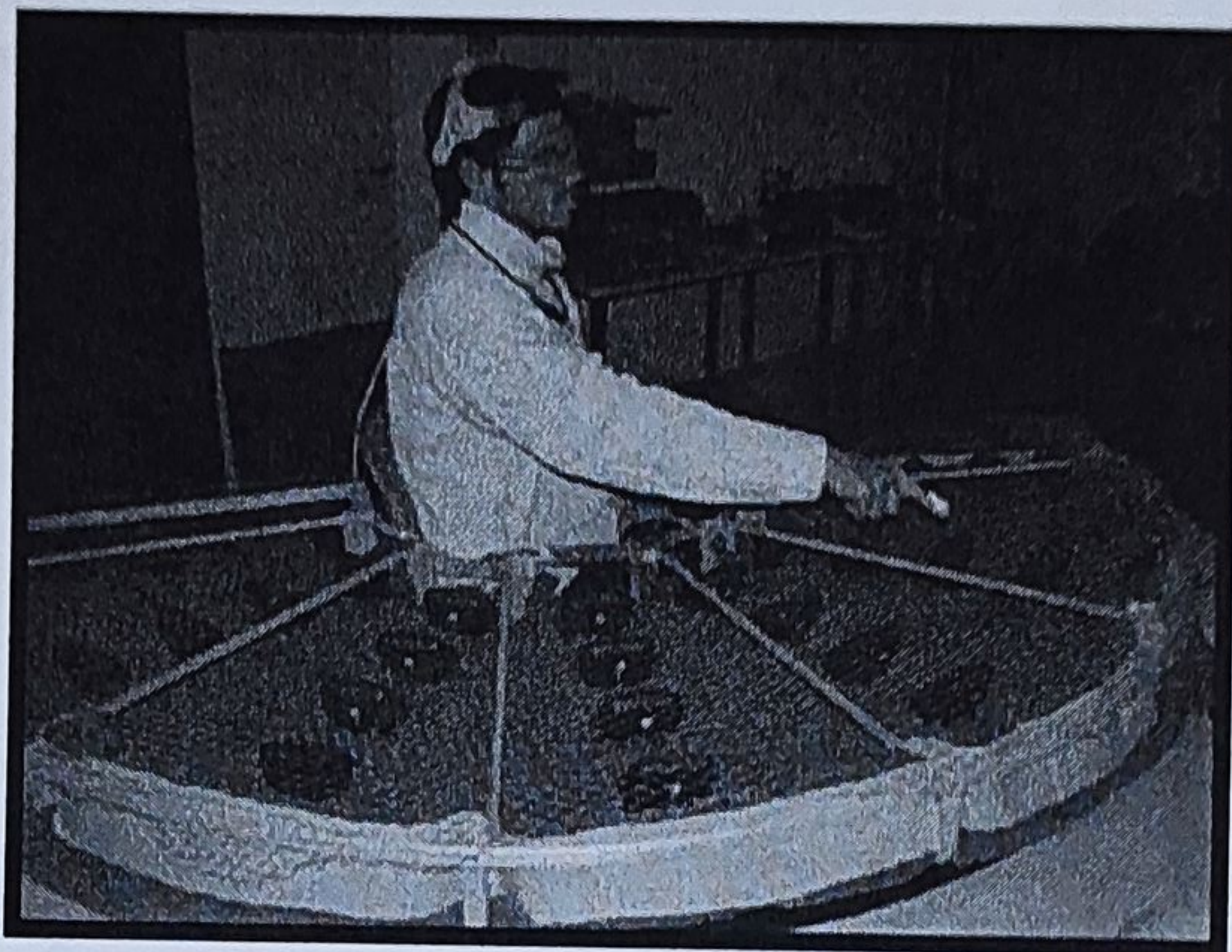


Fig. 2. Dispozitiv experimental cu captorul de poziție pentru cap (fixat pe o cască) și o diodă pe deget care permite urmărirea mișcărilor de marcă (imagine preluată în timpul experimentului)

Mediul permite, în același timp, efectuarea de cercetări în amonte în domeniul psihologiei experimentale (localizare auditivă 3D de exemplu), dar și a unor studii în aval, cu privire la eficiența utilizării dispozitivelor de supleanță.

4. Localizarea rapidă a obiectelor ca ipoteză de supleanță spațială la persoanele cu deficiențe vizuale

4.1 Introducere

Imaginați-vă o clipă că vă aflați într-o cameră întunecată necunoscută. Pronunțați cuvântul „ceașcă” și auziți obiectul respectiv răspunzându-vă printr-un sunet anume. Puteți să vă îndreptați spre ceașcă și să o luați. Sau, pronunțați cuvinte ce descriu mediul și auziți fiecare obiect prezent în câmpul vederii (de ex. fereastra, ușa, scaunele etc.) emițând un sunet adecvat și distinctiv. Acesta este principiul pe care echipa ELIPSE își propune să îl studieze și să îl aplice supleanței spațiale pentru persoanele cu deficiențe vizuale. Ipoteza avansată este că o astfel de abordare poate permite recuperarea unei categorii de comportamente senzorial-motorii simple și funcționale: capacitatea de a se orienta și de a se îndrepta către o țintă, în scopul de a o apuca (e.g. localizarea unei cești într-o încăpere și apucarea acesteia). O altă ipoteză absolut nouă este că localizarea „precisă” a unui număr mic de obiecte ar permite generarea unor reprezentări mentale rare, dar funcționale în spațiul înconjurător (e.g. o încăpere, un loc geografic), și, prin urmare, ameliorarea compor-

CHRISTOPHE JOUFFRAIS

tamentelor senzorial-motorii de nivel mai ridicat, precum navigarea. Este posibil să se înțeleagă organizarea generală a unei camere dacă se cunoaște localizarea ușii, a ferestrelor și a mobilierului principal. Este posibil să se înțeleagă organizarea unei piețe, dacă se cunoaște poziția clădirilor sau a străzilor alăturate și localizarea fântânii.

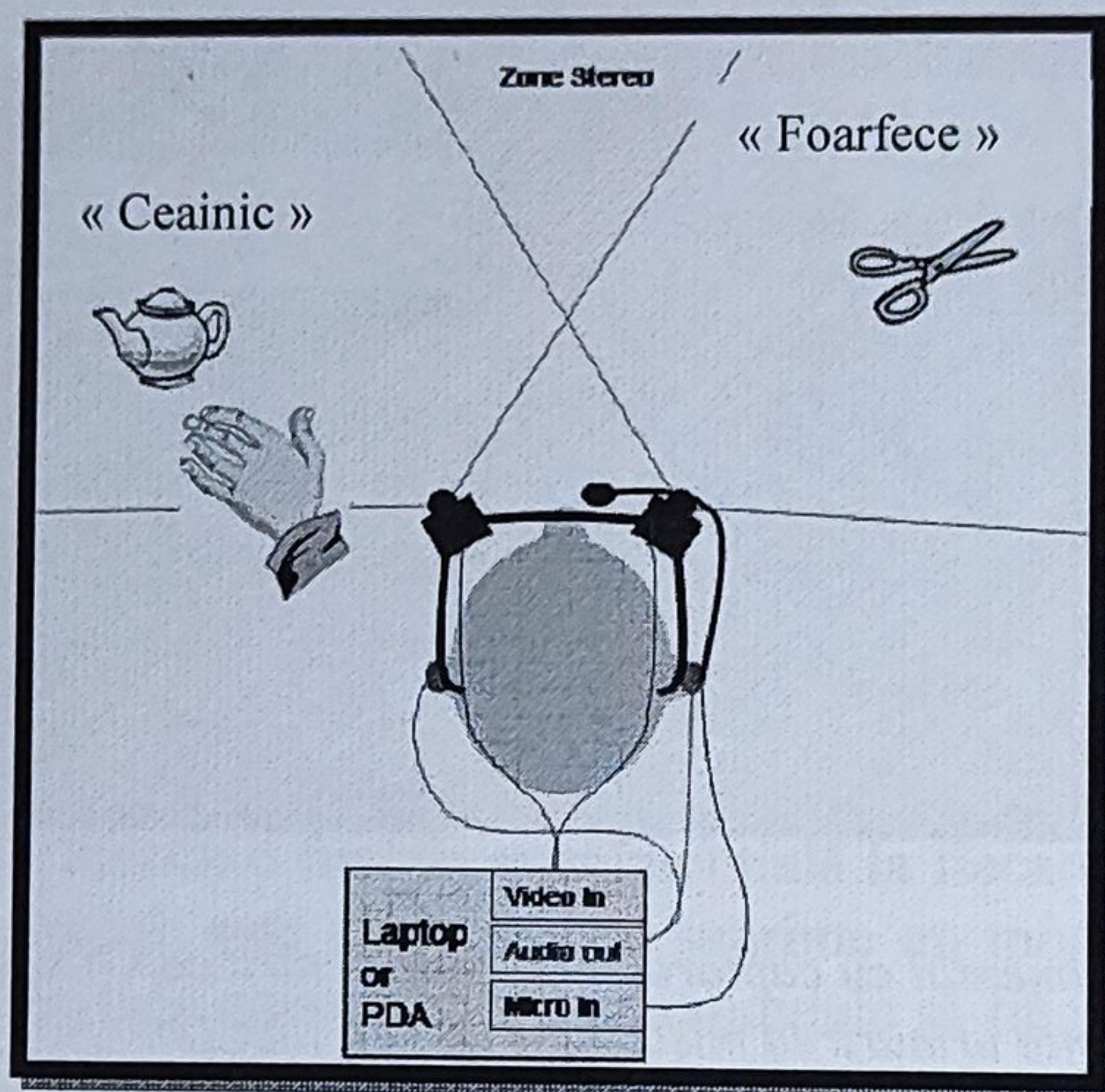


Fig. 3. Schemă care ilustrează principiul supleantei funcționale prin localizarea de obiecte, Principiu comun celor două dispozitive concepute

Figura 3 prezintă o schemă care ilustrează principiul supleantei funcționale prin localizarea de obiecte. Subiectul poartă o pereche de camere legate la un laptop care conține un sistem de recunoaștere și de localizare a obiectelor în timp real. Calculatorul are o intrare pentru microfon cuplată cu un sistem de recunoaștere vocală și o ieșire audio stereofonică. Atunci când utilizatorul vrea să găsească un obiect precis, spune numele obiectului la microfon și primește imediat o informație spațială care îi permite să îl localizeze

4.1.1 Un principiu, mai multe abordări

În cadrul proiectului care a început în 2005, s-au studiat mai multe dezvoltări tehnologice primordiale ale dispozitivului, și anume: vederea artificială care permite recunoașterea și localizarea rapidă a unor obiecte de interes dintr-un anumit spațiu, prototiparea modulară a dispozitivului și două metode de redare a informației spațiale (prin sinteză binaurală și prin interacțiunea creier-mașină). Fig. 4. ilustrează posibilitatea de a utiliza același modul de recunoaștere și localizare pentru a-l conecta cu diferitele metode de redare.

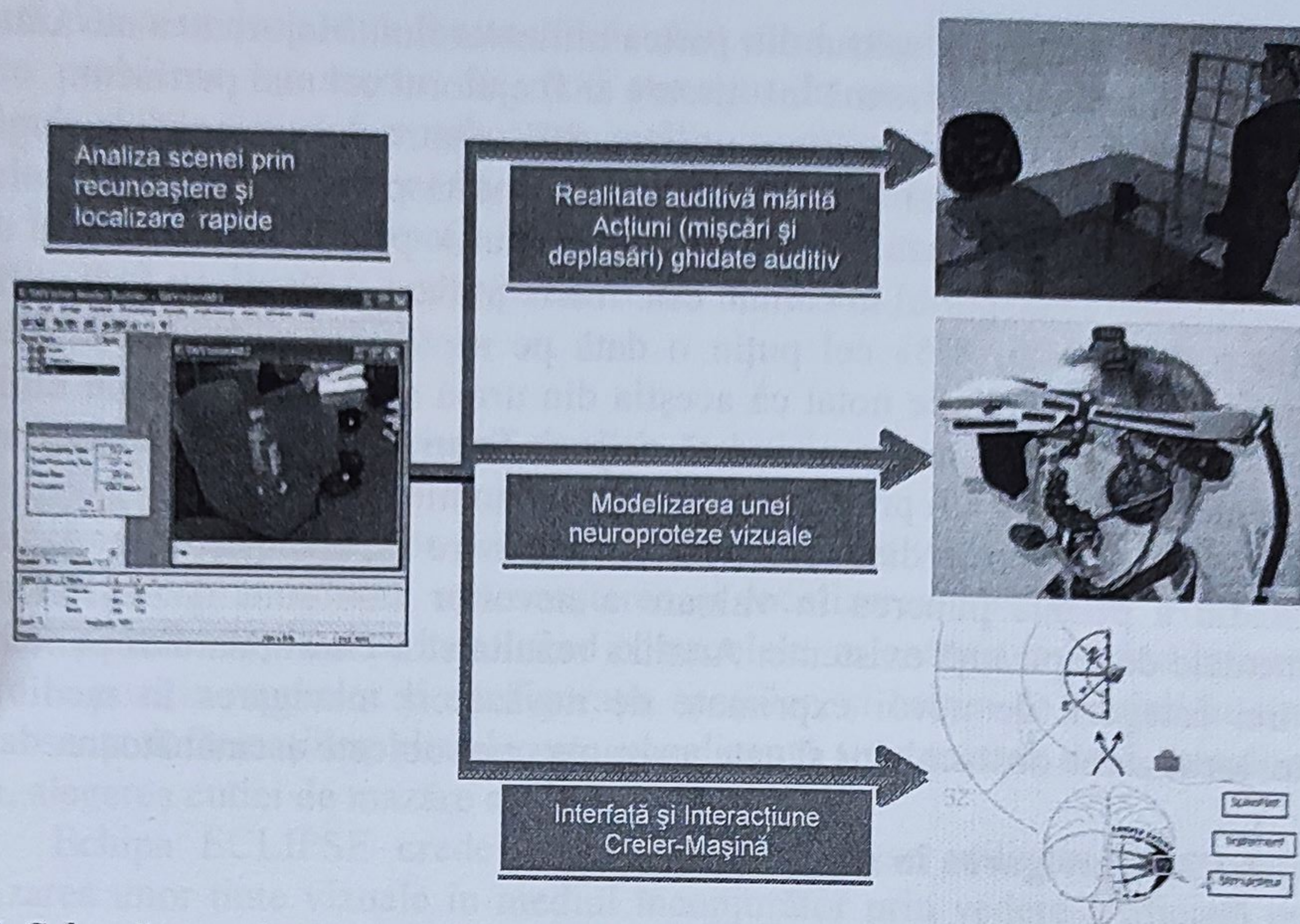


Fig. 4. Schemă care sintetizează diferitele lucrări întreprinse cu privire la supleanța spațială bazată pe localizarea și recunoașterea de obiecte în scenă.

Figura 4 prezintă o schemă care sintetizează diferite lucrări întreprinse pentru supleanța spațială bazată pe localizarea și recunoașterea de obiecte în scenă. Analiza scenei, realizată în timp real⁶⁷, permite selectarea și localizarea (în 3D) a obiectelor de interes. Se dezvoltă trei abordări complementare care permit redarea informațiilor:

- sus: abordare a realității mărite prin sinteză binaurală;
- mijloc: abordare prin simularea unei neuroproteze vizuale;
- jos: abordare reală a unei neuroproteze vizuale la animal.

4.1.2 Analiza deplasării persoanelor cu deficiențe vizuale

În plus față de discuțiile cu IJA, s-a realizat o anchetă pe 54 de nevăzători cu vârste cuprinse între 20 și 63 de ani, prin intermediul unui chestionar electronic accesibil pe internet, care a permis identificarea nevoilor persoanelor cu deficiențe vizuale în timpul deplasărilor lor cotidiene.

O singură persoană utilizează un instrument electronic de ajutor pentru navigare (Teletact), însă menționează un cost prea ridicat în comparație cu utilitatea acestuia. Bastoanele albe sunt utilizate de 76 % dintre subiecții interogați, deoarece permit detectarea obstacolelor apropiate de sol. În schimb, acestea sunt inoperante în cazul obstacolelor a căror înălțime este superioară celei a genunchilor. Principalul lor atu este costul, dar și rapiditatea învățării utilizării lor. 28% dintre subiecții interogați au abandonat bastonul alb în ziua în care au avut un câine ghid, datorită facultății lor de a se adapta, de a învăța noi trasee, de a recunoaște noi puncte de interes și toate

⁶⁷ Se utilizează aici o semnificație personală a timpului real: aceasta corespunde unor procese compatibile cu un comportament uman în curs.

acestea cu un efort cognitiv scăzut din partea utilizatorului. Majoritatea nevăzătorilor interogați consideră că o persoană însoțitoare ar fi ajutorul cel mai pertinent.

Oricare ar fi tipul de asistare utilizat, 80% dintre participanți ies singuri pe stradă cel puțin o dată pe zi, iar 87% cel puțin o dată pe săptămână. Numai 7,5% dintre persoane nu utilizează niciun sistem de ajutor pentru navigare, dar nu ies niciodată singuri. Transportul în comun este foarte utilizat, 38% dintre ei îl utilizează cel puțin o dată pe zi, 85% cel puțin o dată pe săptămână și numai 7,5% nu îl utilizează niciodată. Trebuie notat că aceștia din urmă sunt și cei care nu utilizează niciun sistem de ajutor și nu ies niciodată singuri. Transportul în comun, deși nu este destul de accesibil, pare a fi primordial pentru autonomie.

În spatele acestor date descriptive cu privire la obiceiurile nevăzătorilor, acest studiu a permis punerea în valoare a nevoilor care sunt rar satisfăcute de instrumentele de supleanță existente. Analiza rezultatelor chestionarului pune în evidență trei categorii de nevoi exprimate de nevăzători: navigarea în medii necunoscute, localizarea obstacolelor și categorizarea unor obiecte asemănătoare.

4.1.2.1 Navigarea în medii necunoscute

Majoritatea persoanelor care utilizează un câine pentru orbi în consideră mult mai util decât bastonul alb, deoarece este mult mai ușor de utilizat. Spre deosebire de baston, câinii ghid pot găsi calea cea mai scurtă pentru a ajunge la o destinație, pot să o ia pe trecerile pentru pietoni, pe trotuare și să adapteze un traseu în funcție de obstacolele apărute la distanță sau pe înălțime. În schimb, ei nu sunt acceptați de toate persoanele și, în plus, nu sunt acceptați în toate locurile publice (spitale, magazine...). Prin urmare, câinele este foarte util pentru mobilitate (păstrarea direcției și evitarea obstacolelor) și orientare într-un spațiu imediat (găsirea celei mai apropiate treceri de pietoni). Din păcate, câinele nu poate deloc să genereze un itinerar către o destinație dorită. Stăpânul trebuie să știe, în orice moment, unde se află și în ce direcție vrea să meargă. Aceasta înseamnă să fie corect orientat, să selecționeze itinerariul dorit și să recunoască reperele de navigare în timpul deplasării.

4.1.2.2 Localizarea obstacolelor și a obiectelor

O problemă importantă este detectarea obstacolelor, în special a celor care se află la înălțimea capului (extinctoare, bene de camion etc.), deoarece acestea nu sunt detectate de baston și sunt uneori ignorate de câinii ghid. Formele obstacolelor care sunt cel mai dificil de evitat sunt lungi și fine, verticale sau orizontale. Participanții menționează că obstacolele cele mai periculoase în navigare sunt rupturile brutale de pantă, pe de o parte, și mobilierul urban mobil și vehiculele, pe de altă parte.

4.1.2.3 Categorizarea obiectelor asemănătoare

A recunoaște un obiect prin forma sa poate să nu fie suficient atunci când obiectele sunt apropiate din punct de vedere structural. De exemplu, este util de diferențiat o cutie de mazăre de una de ananas sau o factură de chirie de o chitanță

bancară. De asemenea, este foarte util de știut cărui magazin îi aparține ușa detectată. Aceste probleme au revenit în mod recurent în ancheta realizată și subliniază problema dobândirii informației semantice cu privire la un obiect (de exemplu numele străzii, valoarea unei bancnote, numărul autobuzului, numele unui magazin etc.). În lipsa unei informații semantice complete, culoarea (semafor, culoarea hainelor etc.) și luminozitatea (vremea de afară etc.) sunt informații care se pot dovedi indispensabile.

4.1.2.4 Concluzii privind analiza

Analiza răspunsurilor persoanelor cu deficiențe vizuale scoate la iveală două funcții vizuale primordiale: recunoașterea și localizarea obiectelor. Această dublă facultate permite deplasarea localizând obstacolele, orientarea culegând informații cu privire la poziția reperelor și configurația spațiilor, și luarea de decizii în situații în care absența informațiilor vizuale este penalizantă (traversarea dacă semaforul este verde, alegerea cutiei de mazăre etc.).

Echipa ECLIPSE crede că este posibil să se utilizeze recunoașterea și localizarea unor ținte vizuale în mediul înconjurător prin vedere artificială pentru a ajuta persoanele nevăzătoare să navigheze. De exemplu, s-ar putea implementa o funcție de recunoaștere a unor repere vizuale (metrou, ușa, trecere de pietoni, logo etc.). Această informație ar permite păstrarea unei direcții, îndreptarea către o destinație și localizarea punctelor de reper sau a punctelor de interes, cum ar fi clădirile importante. Faptul de a recupera această funcție a sistemului vizual ar permite să se răspundă, în același timp, unei alte nevoi, cea a recunoașterii obiectelor cu forme asemănătoare, dar pentru care numai motivul vizual este diferit.

4.2 Un modul de recunoaștere și de localizare rapidă a obiectelor

Astăzi, mulți algoritmi care privesc vederea ajung la rezultate de segmentare și de interpretare a scenei foarte evaluate. Din contră, foarte puțini dintre aceștia sunt destul de rapizi pentru a fi compatibili cu un comportament în curs, de exemplu pentru a asista o persoană cu deficiențe vizuale într-o sarcină de navigare sau de apucare a unor obiecte. În colaborare cu Simon J. Thorpe, s-a arătat că este posibil să se utilizeze un sistem de recunoaștere a obiectelor (SpikeNet) pentru a localiza cu precizie obiectele în spațiu. Acest sistem, ca urmare a timpului de calcul foarte scurt (câteva zeci de ms) și a rezistenței sale, este utilizabil, în același timp, pentru navigare și pentru apucarea obiectelor (a se vedea Florian Dramas⁶⁸ et al. 2007; Florian Dramas⁶⁹ et al. 2010; Florian Dramas⁷⁰ 2010).

⁶⁸ Dramas, Florian, Simon J. Thorpe & Christophe Jouffrais, *Localisation d'objets pour les non-voyants par analyse d'image: analyse du besoin et prototypage*, Paper presented to the ASSISTH, Toulouse, FR, 2007.

⁶⁹ Dramas, Florian, Simon J. Thorpe & Christophe Jouffrais, *Artificial Vision For The Blind: A Bio-Inspired Algorithm For Objects And Obstacles Detection*, International Journal of Image and Graphics 10.531-44, 2010.

⁷⁰ Dramas, Florian, *Localisation d'objets pour les non-voyants: augmentation sensorielle et neuroprothèse*, PhD: University of Toulouse 3, France, 2010.

În strânsă colaborare cu CerCo (echipa lui Simon J. Thorpe), s-au dezvoltat mai multe scenarii predefinite de supleanță pentru nevăzători pentru care se optimizează SpikeNet și se compară cu alte softuri de localizare a obiectelor.

Se încearcă dezvoltarea unui algoritm care să poată fi generalizabil (nu numai într-un mediu cunoscut), care să poată să urmărească obiectele și să reziste la rotații importante. O altă problemă importantă la care se lucrează în prezent este crearea de modele, care, în prezent, este supervizată și trebuie să treacă la un mod semi-supervizat.

4.3 Asistarea comportamentelor ghidate vizual și a cogniției spațiale

4.3.1 Localizarea și apucarea unor obiecte, ghidate prin sinteză binaurală

Localizarea auditivă umană este bazată pe indicii binaurale (diferențe interaurale de timp și de intensitate) și pe indicii spectrale (monaurale) generate de particularitățile fizice ale observatorului, în special forma capului și a pavilioanelor (Middlebrooks & Green⁷¹ 1991). Toate aceste indicii spațiale auditive pot fi încorporate în funcții de transfer care depind de cap (Head-Related Transfer Function, HRTF), care sunt măsurate în urechea subiectului ca răspuns la sursele sonore din spațiu (Kulkarni & Colburn⁷² 1998). În consecință, ascultarea unor stimuli sonori corecți cu o HRTF permite menținerea informației spațiale proprii subiectului și reproducerea unei percepții tridimensionale.

Experiențele făcute pe transformările auditivo-motrice arată că mișcările îndreptate către o țintă auditivă sunt posibile și precise atât la subiecții normali, cât și la nevăzători (Zwiers⁷³ et al. 2001a; Zwiers⁷⁴ et al. 2001b; Fujiki⁷⁵ et al. 2002). Unul dintre modulele noastre de restituire are la bază sinteza binaurală (sinteză a sunetelor virtuale în 3D) ca mijloc de localizare spațială a obiectelor căutate, cu ipoteza că este posibilă localizarea și apucarea unui obiect marit printr-un sunet 3D virtual.

Pentru a modeliza și a reproduce prin sinteză binaurală localizarea spațială auditivă, împreună cu Brian F.G. Katz (LIMSI, Orsay) s-a încercat: 1/ înțelegerea capacităților perceptive comparate ale văzătorilor și ale nevăzătorilor și 2/ conceperea unei metode de interacțiune sonoră 3D care permite localizarea obiectelor spațiale.

⁷¹ Middlebrooks, John C. & David M. Green, *Sound localization by human listeners*, Annu.Rev.Psychol, 42,135-59, 1991.

⁷² Kulkarni, Abhijit & H. Steven Colburn, *Role of spectral detail in sound-source localization*, Nature 396,747-49, 1998.

⁷³ Zwiers, Marcel P., John A. Van Opstal & Johan R. Cruysberg, *A spatial hearing deficit in early-blind humans*, Journal of Neuroscience 21,RC142-RC45, 2001a.

⁷⁴ Zwiers, Marcel P., John A. Van Opstal & Johan R. Cruysberg, *Two-dimensional sound-localization behavior of early-blind humans*, Experimental Brain Research 140,206-22, 2001b.

⁷⁵ Fujiki, Nobuya, Klaus A. Riederer, Veikko Jousmaki, Jyrki P. Makela & Riitta Hari, *Human cortical representation of virtual auditory space: differences between sound azimuth and elevation*, Eur.J.Neurosci. 16,2207-13, 2002.

S-a dezvoltat un post experimental (platformă cu 35 de difuzoare așezate în arc de cerc în spațiul peripersonal) care permite să se prezinte sunete reale în spațiul din jurul subiectului, precum și să se urmărească și să se analizeze traiectoria mâinii. Într-o primă experiență (a se vedea Florian Dramas⁷⁶ et al. 2008), s-au înregistrat mișcările de punctare a 19 subiecți văzători și nevzăzători către diferite tipuri de sunete reale (durată și repetiții variabile).

Figura 6 prezintă utilizarea instrumentului COFT. Cercurile verzi corespund poziției fiecărui difuzor, în reperul imagine. Primul cerc roșu corespunde poziției estimate în timp real a indexului utilizatorului. Al doilea cerc roșu corespunde difuzorului de la care provine stimulul. Înregistrările mișcărilor de punctare a mâinii către sursa sonoră au fost realizate la IRIT în 2D prin captură video. Acest post experimental este compatibil cu sistemul de captură a mișcării 3D NaturalPoint.

Erorile de azimut arată că precizia marcată este mai bună în fața subiectului și scade în lateral, pentru a deveni rea în spatele subiectului. Mai precis, se arată că precizia localizării în fața subiectului este compatibilă cu un gest de atingere și de apucare (o eroare de 6° reprezintă o distanță de 6 cm până la 57 cm). Erorile în ceea ce privește distanța sunt puțin mai importante în față (de ordinul a 12 cm) decât în lateral (de ordinul a 10 cm). Pe ansamblu, aceste rezultate arată că dacă poziția unui obiect ar fi desemnată printr-un sunet real, ar fi posibil ca mâna să se apropie suficient de obiect pentru a-l apuca.



Fig. 5. Dispozitiv experimental cu captor de poziție pentru cap (fixat pe o cască) și dioda pe deget care permite urmărirea mișcărilor de marcare.

Într-o a doua experiență, s-a cerut subiecților să efectueze mișcări de marcare către un zgomot alb cu durate și repetări diferite. Eroarea de azimut în funcție de experimente arată că precizia de localizare angulară este dependentă de durata stimulului și de numărul de repetări ale sunetului.

⁷⁶ Dramas, Florian, Brian F.G. Katz & Christophe Jouffrais, *Auditory-guided reaching movements in the peripersonal frontal space*, Paper presented at the The Journal of the Acoustical Society of America, 2008.

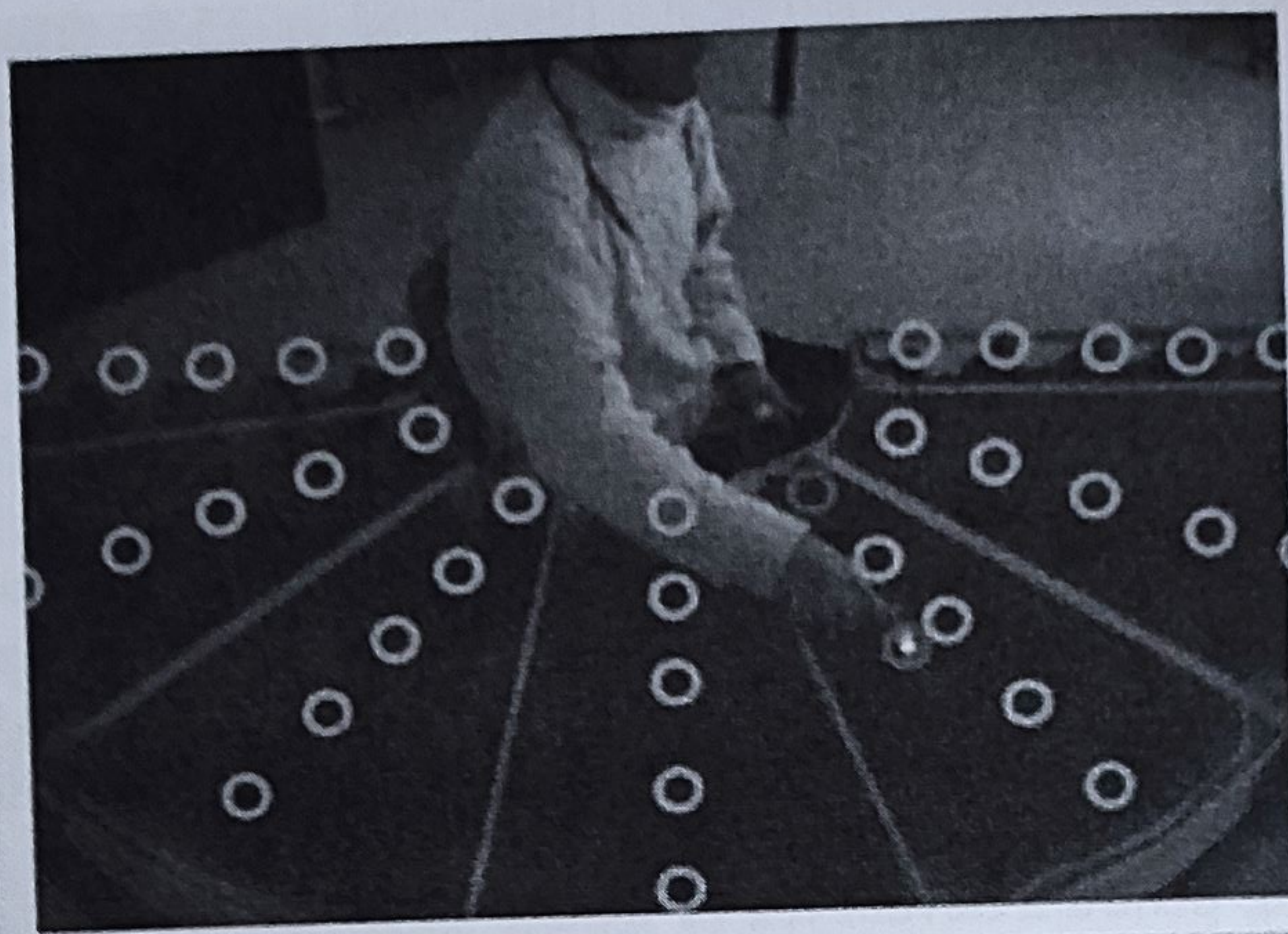


Fig. 6. Utilizarea instrumentului COFT.

Se arată, în special, că precizia nu se ameliorează la nesfârșit cu numărul de sunete și cu durata sunetelor și că precizia este deja suficientă și pare să atingă un anumit grad pentru două sunete de 25 ms întrerupte de 30 ms de pauză.

Paralel cu aceste lucrări privind localizarea unor sunete reale, institutul LIMSI a dezvoltat un motor de sinteză (LSE, LIMSI Spatialization Engine) care va permite compararea acțiunilor ghidate de sunete reale și virtuale și lucrează asupra designului sonor.

Acest motor va permite compararea răspunsurilor subiecților nevăzători cu cele ale subiecților văzători lipsiți de vedere și, prin urmare, cunoașterea capacităților de percepție, de reprezentare spațială și de acțiune a subiecților nevăzători față de cele ale subiecților văzători, indiferent dacă ținta auditivă este reală (difuzoare plasate în apropierea obiectului) sau virtuală (realitate mărită prin sinteză binaurală).

El permite, de asemenea, conceperea unei interfețe sonore 3D care poate să ghideze mâna sau corpul către o țintă auditivă și să evalueze posibilitățile de transfer ale unor atribute suplimentare privind obiectele (orientarea, de exemplu). Echipa lui Brian F.G. Katz (a se vedea Brian F.G. Katz & Picinali⁷⁷ 2011) a arătat mai ales că persoanele cu deficiențe vizuale sunt capabile să înțeleagă un mediu sonor constituit din mai multe surse virtuale localizate (Afonso⁷⁸ et al. 2005).

4.3.2 Asistarea navigării bazate pe localizarea țintelor

Localizarea unor ținte în mediul înconjurător apropiat permite utilizatorului să restaureze unele acțiuni ghidate vizual, în special mișcarea de orientare către o țintă cuprinsă în câmpul vederii sau mișcarea de apucare a unui obiect prezent în

⁷⁷ Katz, Brian F. G. & Lorenzo Picinali, *Spatial Audio Applied to Research with the Blind*, Advances in Sound Localization, ed. by P. Strumillo: InTech, 2011.

⁷⁸ Afonso, Amandine, Brian F.G. Katz, Alan Blum & Michel Denis, *Mental imagery and the acquisition of spatial knowledge without vision: A study of blind and sighted people in an immersive audio virtual environment*, St Andrews, Scotland, 2005.

câmpul peripersonal. Totuși, această asistare nu este suficientă pentru a permite unor persoane cu deficiențe vizuale să navigheze în spații largi. Ajutoarele electronice pentru orientare (EOA) furnizează utilizatorilor informații cu privire la poziția lor și la direcțiile pe care trebuie să le urmeze. Un EOA este compus din trei elemente indispensabile: 1/ un modul de poziționare bazat pe semnalele satelitare; 2/ un sistem informațional geografic (SIG) cu o bază de date spațiale, un soft pentru selectarea de itinerare și urmărirea utilizatorilor; și 3/ o interfață utilizator care are la bază o interacțiune nevizuală (de exemplu, vocală sau tactilă).

Până în prezent, aceste EOA sunt bazate, în principal, pe GPS și pe serviciile bazate pe localizare (LBS). Anumite aparate comerciale sunt disponibile (BrailleNote, Angeo, Trekker), dar, în majoritatea cazurilor, utilizarea lor a fost limitată de prețul relativ ridicat (aproximativ 2000 \$) și de o proastă precizie a poziției (uneori mai mare de 10 m, în special în zonele urbane). Ședințele de analiză cu utilizatorii și educatorii în domeniul orientării și mobilității (O&M) au permis descoperirea unei alte limitări importante. Se pare că aceste sisteme au la bază hărți comerciale concepute pentru navigarea vehiculelor. Poziționarea și ghidarea propuse în sistemele comerciale se sprijină deci pe o rețea de străzi și nu pe o rețea de drumuri pentru pietoni. Evident, toate reperele necesare pentru navigarea persoanelor cu deficiențe vizuale (puncte de reper, puncte de confirmare și puncte de interes) sunt absente din baza de date. Ceea ce se propune aici este că această abordare de localizare a țintelor poate permite ameliorarea poziției și, prin urmare, ajutarea persoanelor cu deficiențe vizuale să navigheze, dar trebuie totuși să se sprijine pe un SIG adaptat.

Dispozitivul de ajutor pentru navigare conceput este bazat pe prototipul de ajutor pentru localizarea țintelor, care permite evaluarea unei asistări pentru navigare adaptată în special persoanelor cu deficiențe vizuale. Până în prezent, s-a lucrat pe două verigi majore ale sistemelor actuale: sistemul informațional geografic (SIG) și poziționare. Dispozitivul complet este în curs de evaluare⁷⁹.

4.3.2.1 Un sistem informațional geografic adaptat pietonilor cu deficiențe vizuale

SIG poate fi definit ca un instrument de captură, manipulare, afișaj, interogare și analiză a datelor spațiale (a se vedea Burrough⁸⁰ 1994). SIG cuprind o bază de date digitală și softuri concepute pentru a selecționa itinerare, pentru a urmări traseul călătorului și a informa utilizatorul cu privire la caracteristicile locale.

S-a făcut o clasificare a obiectelor care trebuie să fie integrate în baza de date a SIG. S-au definit patru clase principale incluzând Punctele de Interes (POI), indiciile și punctele de reper (LM), zonele pietonale (sidewalks) și țintele vizuale (VP). POI reprezintă locuri sau obiecte care sunt destinații potențiale. Acestea sunt utile și pentru a permite pietonului să înțeleagă mai bine mediul în care se deplasează

⁷⁹ Aceste lucrări fac parte din proiectul NAVIG, aprobat de polul AESE și cofinanțat de ANR (programul TecSan 2008) și Regiunea Midi-Pyrénées (programul APRRTT 2008). A se vedea <http://navig.irit.fr/>

⁸⁰ Burrough, Peter A., *Principles of geographical information systems for land resources assessment*, Clarendon Press, 1994.

(de exemplu clădirile publice, magazinele, serviciile etc.). Un indiciu, după cum arată și numele său, desemnează un loc care îi permite persoanei cu deficiențe vizuale să facă inferențe cu privire la propria sa poziție. Acesta este, în esență, nevizual. Un punct de reper corespunde unei acumulări de indicii. Pentru a concepe un SIG adaptat navigării pietonilor, este important să se includă toate drumurile pedestre posibile, ca de exemplu trotuarele și trecerile pentru pietoni. În sfârșit, punctele vizuale (VP) desemnează obiecte geolocalizate care pot fi detectate prin modulul de recunoaștere și de localizare a obiectelor. Acestea au două funcții: localizarea unor obiecte specifice din câmpul vizual (de exemplu, o cutie poștală la care utilizatorul vrea să ajungă) și ajutorul pentru poziționare datorită estimării distanței și direcției între obiectul recunoscut și utilizator.

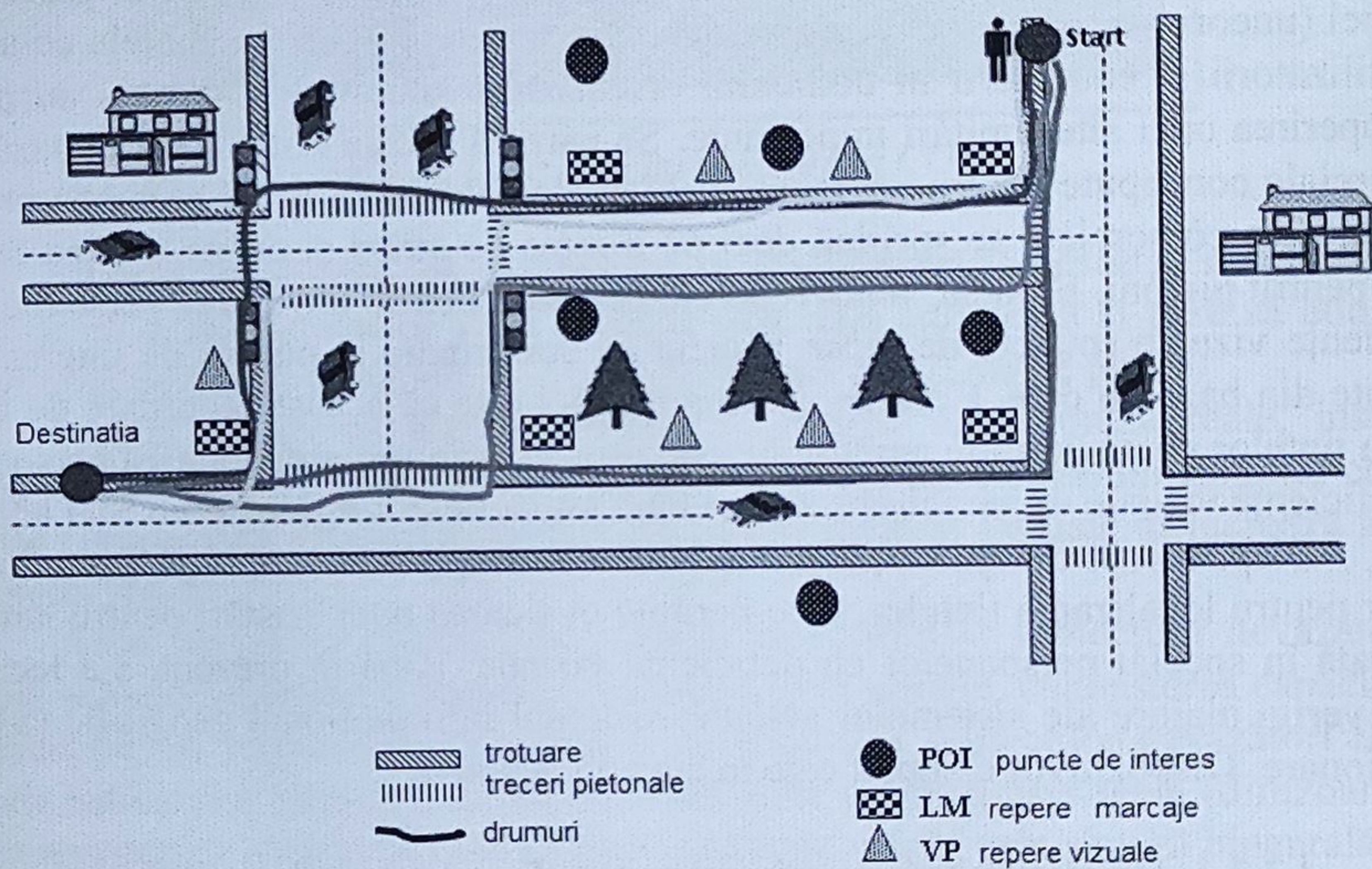


Fig. 7. Ilustrarea a patru itinerare posibile între un punct de plecare și o destinație. (imagine preluată din lucrarea Slim Kammoun⁸¹ et al. 2010)

Figura 7 ilustrează patru itinerare posibile între un punct de plecare și o destinație. Pentru fiecare itinerar se calculează un scor. Calea pietonală care conține POI, LM și VP și care are cel mai bun scor este cea aleasă.

S-a realizat un SIG al campusului Universității din Toulouse, extras din baza de date OpenStreetMap, la care s-au adăugat clasele de obiecte menționate, precum și un algoritm de selecție a itinerarului care include nevoile și preferințele pietonilor cu deficiențe vizuale. Figura 7 arată un exemplu de alegere a unui itinerar problematic. Pentru a ajunge la aceeași destinație, pietonii pot alege unul dintre cele patru itinerare reprezentate prin culori. S-a arătat că traseul optim poate fi selectat rezolvând

⁸¹ Kammoun, Slim, Dramas Florian, Bernard Oriola & Christophe Jouffrais, *Route selection algorithm for Blind pedestrian*, Paper presented to the International Conference on Control, Automation and Systems, Gyeonggi-do, South Korea, 2010.

problema minimizării prin algoritmul lui Dijkstra (Dijkstra⁸² 1959), luând în considerare clasele de obiecte definite (Slim Kammoun⁸³ et al. 2010). Pentru fiecare secțiune, beneficiile și costurile au fost alese pe baza nevoilor și a comportamentului persoanelor cu deficiențe vizuale în timpul realizării anumitor sarcini de navigare. Această metodă selectează, în final, itinerariile care corespund cel mai bine navigării persoanelor cu deficiențe vizuale.

Primul tip de particularitate corespunde zonelor pietonale adaptate (de exemplu, un trotuar larg care permite deplasarea cu un baston sau un câine). Al doilea tip de particularitate constă în a găsi repere nevizuale (LM) care îi ajută pe utilizatori să își confirme propria poziție. Această alegere a itinerariului ar trebui să fie însoțită de o ghidare adaptată care permite semnalarea LM și a POI întâlnite pe parcursul traseului. O astfel de ghidare ar avea două consecințe potențiale importante: ar minimiza stresul generat de deplasări și ar permite dobândirea, în plus față de competențele procedurale egocentrice, a unor cunoștințe spațiale alocentrice privind configurația mediului înconjurător. Evident, o ghidare precisă are la bază o poziționare precisă.

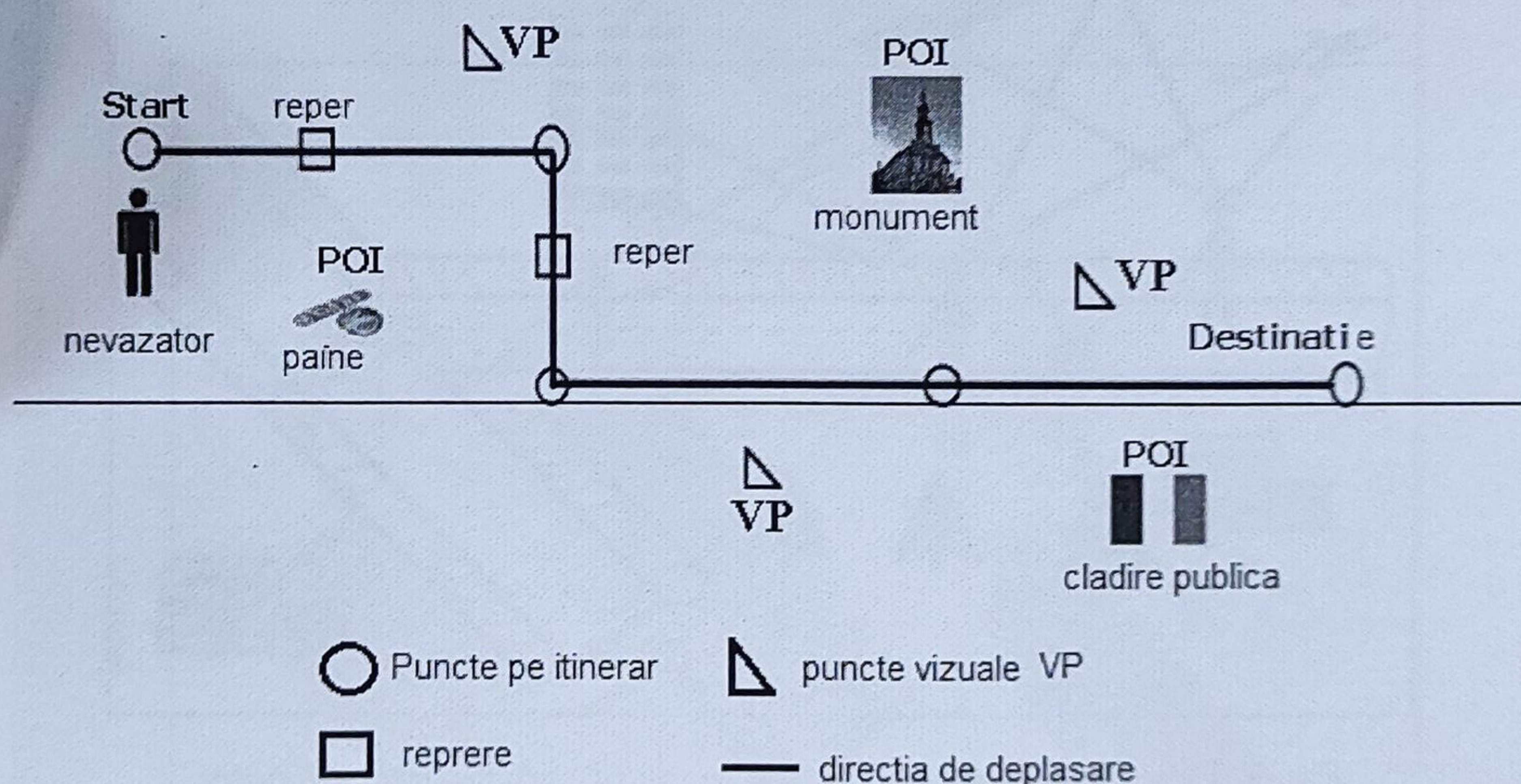


Fig. 8. Ilustrarea unui itinerar selectat care include diferite clase de obiecte. „Landmarks”-urile îi permit utilizatorului să își confirme poziția în itinerar. VP pentru localizarea obiectelor de interes în câmpul vizual (imagine preluată din Slim Kammoun et al, 2010).

Figura 8 ilustrează un itinerar selectat care include diferite clase de obiecte. „Landmarks”-urile îi permit utilizatorului să își confirme poziția în itinerar. VP sunt utile pentru a localiza obiecte de interes în câmpul vizual. Aceste puncte și POI permit dobândirea unor cunoștințe alocentrice cu privire la mediul parcurs.

⁸² Dijkstra, Edsger W., *A note on two problems in connexion with graphs*, Numerische mathematik 1.269-71, 1959.

⁸³ Kammoun, Slim, Dramas Florian, Bernard Oriola & Christophe Jouffrais, *Route selection algorithm for Blind pedestrian*, Paper presented to the International Conference on Control, Automation and Systems, Gyeonggi-do, South Korea, 2010.

4.3.2.2 Fuziunea datelor cu privire la poziționare

Precizia poziționării în numeroase medii, cum ar fi orașele, este rareori mai bună de 10–20 de metri. În aceste condiții, EOA nu sunt cu adevărat utilizabile, deoarece conduc la situații periculoase (de exemplu, traversarea în afara zonelor protejate) sau la erori prejudiciabile (de exemplu, a lua pe o stradă greșită). În scopul ameliorării poziționării, s-au făcut numeroase propuneri. Diferite proiecte de cercetare (Helal⁸⁴ et al. 2001; Petrie⁸⁵ et al. 1996; Loomis⁸⁶ et al. 1994) recomandă să se utilizeze un GPS diferențial (DGPS) care reduce marja de eroare nominală la mai puțin de 1 metru în condiții ideale. Cu toate acestea, DGPS are la bază utilizarea unei rețele fixe de stații de referință la sol. Această rețea este relativ costisitoare și nu este foarte extinsă. În plus, greutatea și mărimea receptorului (cel puțin 0,5 kg) nu sunt adecvate pentru mobilitatea pietonilor. Una dintre soluții constă în utilizarea unor indicii dinamice în ceea ce privește deplasarea utilizatorului.

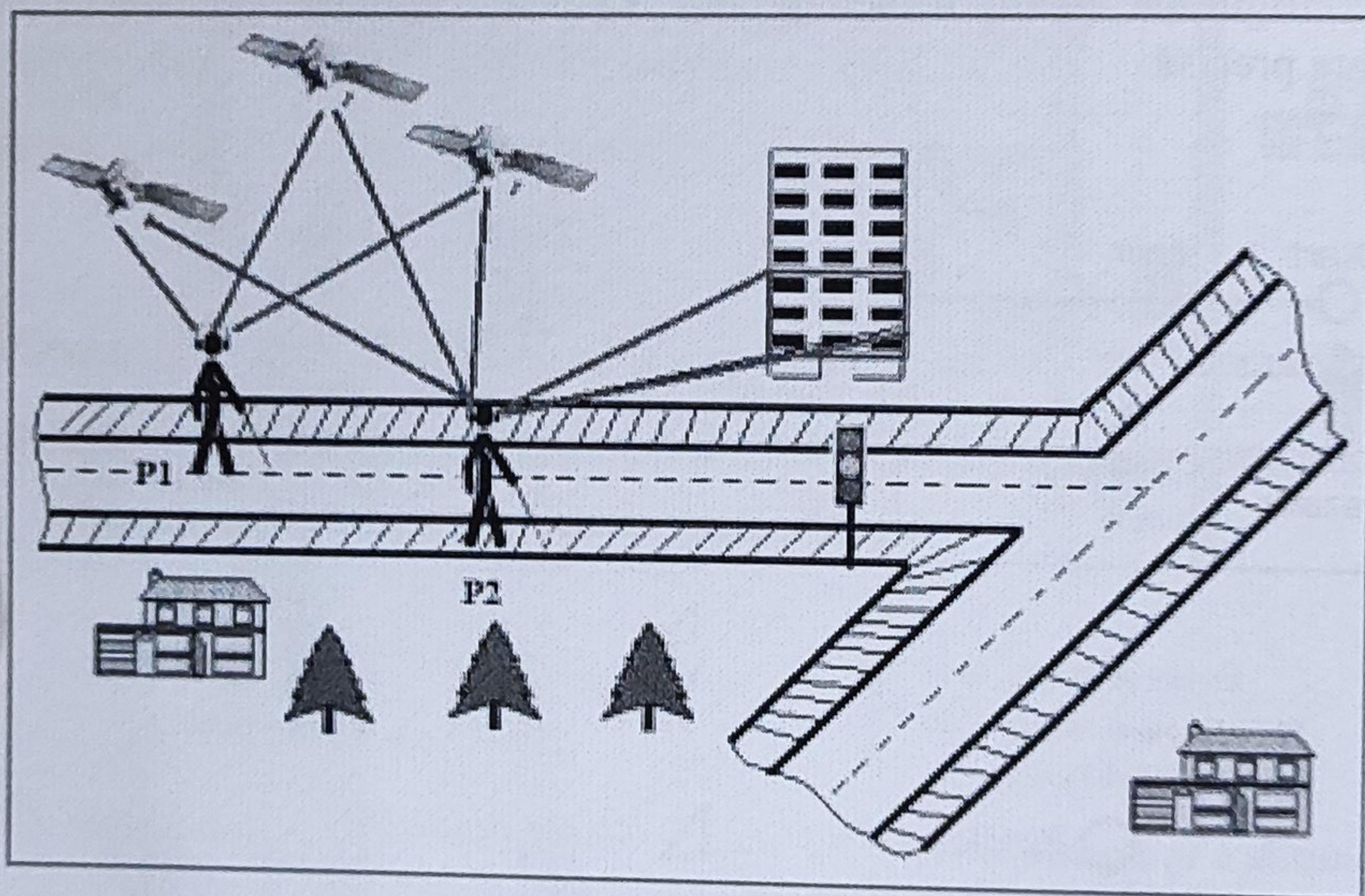


Fig. 9. Principiul de funcționare al modului „poziționare utilizator”
(figură preluată cu acordul autorului din lucrarea Brilhault et al 2011)

S-a propus ameliorarea poziționării datorită detectării unor indicii vizuale statice prin intermediul modului de vedere integrat (Adrien Brilhault⁸⁷ et al. 2011).

⁸⁴ Helal, Abdelsalam, Steven Edwin Moore & Balaji Ramachandran, *Drishti: An integrated navigation system for visually impaired and disabled*, Paper presented at the ISWC'01, Zurich, Switzerland, 2001.

⁸⁵ Petrie, Helen, Valerie Johnson, Thomas Strothotte, Andreas Raab, Steffi Fritz & Rainer Michel, *MoBIC: Designing a travel aid for blind and elderly people*, Journal of Navigation 49.45-52, 1996.

⁸⁶ Loomis, Jack M., Reginald G. Golledge, Roberta L. Klatzky, Jon M. Speigle & Jerome Tietz, *Personal guidance system for the visually impaired*, Marina Del Rey, California, United States, 1994.

⁸⁷ Brilhault, Adrien, Slim Kammoun, Olivier Gutierrez, Philippe Truillet & Jouffrais Christophe, *Fusion of Artificial Vision and GPS to Improve Blind Pedestrian Positioning*, Paper presented at the 4th IFIP International Conference on New Technologies, Mobility and Security, Paris, France, 2011.

În această funcție, numită „poziționare utilizator”, vederea integrată este utilizată pentru a detecta ținte vizuale care nu sunt semnalate utilizatorului, ci sunt utilizate doar ca puncte de ancorare pentru a preciza poziționarea. În acest caz, nu utilizatorul este cel care determină ce ținte trebuie detectate. Procesul este automat și are la bază localizarea aproximativă a utilizatorului estimată de GPS. Această selectare dinamică și contextuală permite controlarea numărului total de modele active, ceea ce este compatibil cu o funcționare în timp real. Într-adevăr, algoritmul vederii nu caută decât țintele prezente în cartierul în care se situează utilizatorul (a se vedea Fig. 9). S-a verificat experimental că performanțele sistemului sunt compatibile cu utilizarea în timp real. Pe prototipul utilizat (laptop cu processor i7 Intel), s-a putut menține o cadență de recunoaștere la 15Hz utilizând imagini de 320x240 pixeli și s-au încărcat aproximativ 1000 de modele diferite cu o mărime de 180 pixeli. Estimarea poziției utilizatorului datorită detectării unor ținte vizuale situate între 2 și 10 metri distanță este mai mică de 40 cm. Actualmente, se lucrează la evaluarea dispozitivului complet (GPS/vedere integrată/captoare inerțiale) pe un parcurs de aproximativ 500m care posedă 9 ținte geolocalizate (Jiri Borovec⁸⁸ 2011). Rezultatele preliminare indică faptul că această metodă permite suprimarea erorilor comise de GPS (Fig. 10).



Fig. 10. Fuziunea de date heterogene pentru ameliorarea poziționării.

Figura 10 prezintă o fuziune de date heterogene, permițând ameliorarea poziționării. Acest desen reprezintă o hartă a campusului universității cu clădiri (culoarea

⁸⁸ Borovec, Jiri, *Fusion of heterogeneous data for better positioning of visually impaired pedestrians*, Master degree: Université de Toulouse, 2011.

gri) și un itinerar (puncte violet). Linia galbenă corespunde poziționării GPS a unui utilizator pe traseu. Linia albastră corespunde poziționării bazate pe fuziunea datelor GPS cu date de la centrala inerțială și cu detectări ale unor ținte vizuale geolocalizate (stele verzi).

4.3.3 Localizarea și apucarea unor obiecte, ghidate de o interacțiune creier-mașină

Este posibil să se conecteze o cameră la creier prin intermediul unei interfețe neurale. Această interfață este constituită dintr-o matrice care conține maxim o sută de electrozi de stimulare. Ca urmare a acestui număr limitat de electrozi, abordarea clasică de tipul „scoreboard” (un pixel corespunde unui electrod de stimulare) nu este funcțională în prezent.

Pentru validarea abordării de localizare a obiectelor în cadrul unei neuroproteze vizuale funcționale, se lucrează la stimularea unei neuroproteze vizuale într-o cască de realitate virtuală. Această abordare va permite demonstrarea conceptului neuroprotezei funcționale prin localizare și testarea a două ipoteze care privesc recuperarea unor comportamente ghidate vizual și crearea unei hărți cognitive pe baza localizării de obiecte.

Primele etape ale proiectului de simulare au constatat în dezvoltarea unui post experimental care permite plasarea unui subiect într-o situație echivalentă celei a unui nevăzător implantat cu o neuroproteză vizuală.

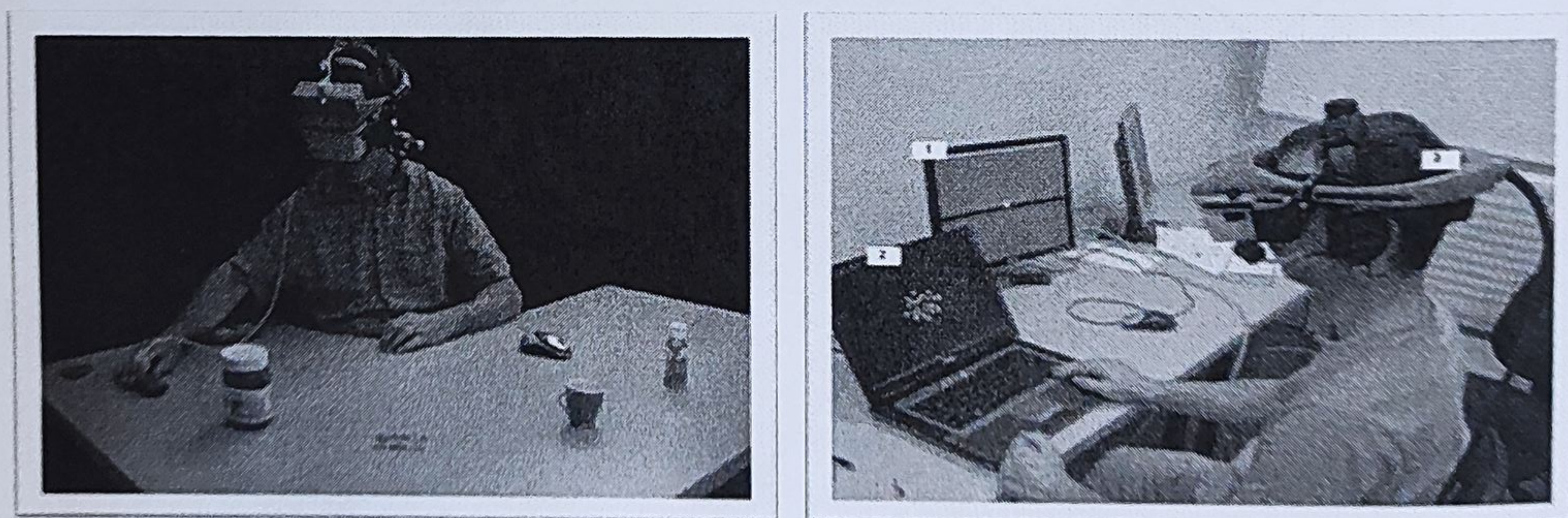


Fig. 11. *Post experimental de stimularea neuroprotezei vizuale: subiectul percepe fofene în cască.*

La stânga, subiectul echipat cu cască și cameră stereoscopică: sarcina sa este identificarea, apucarea diferitelor obiecte de pe o masă.

La dreapta, subiectul este implicat într-o sarcină de navigare într-un spațiu 3D virtual: sarcina sa este de a atinge o țintă într-un labirint virtual.

Fig. 11 reprezintă un subiect care poartă o cască de realitate virtuală. Casca are o funcție dublă: 1/ să priveze subiectul de vederea scenei vizuale, și 2/ să îi transmită informații spațiale numai sub forma unor fofene. Figura prezintă un post experimental de stimulare a unei neuroproteze vizuale. La stânga, subiectul este echipat cu casca și cu o cameră stereoscopică. Sarcina sa este de a identifica, apoi de a apuca diferite obiecte de pe o masă. La dreapta, subiectul este implicat într-o sar-

cină de navigare într-un spațiu 3D virtual. Sarcina sa constă în a atinge o țintă într-un labirint virtual. În ambele cazuri, subiectul nu percepe decât fosfene în casă.

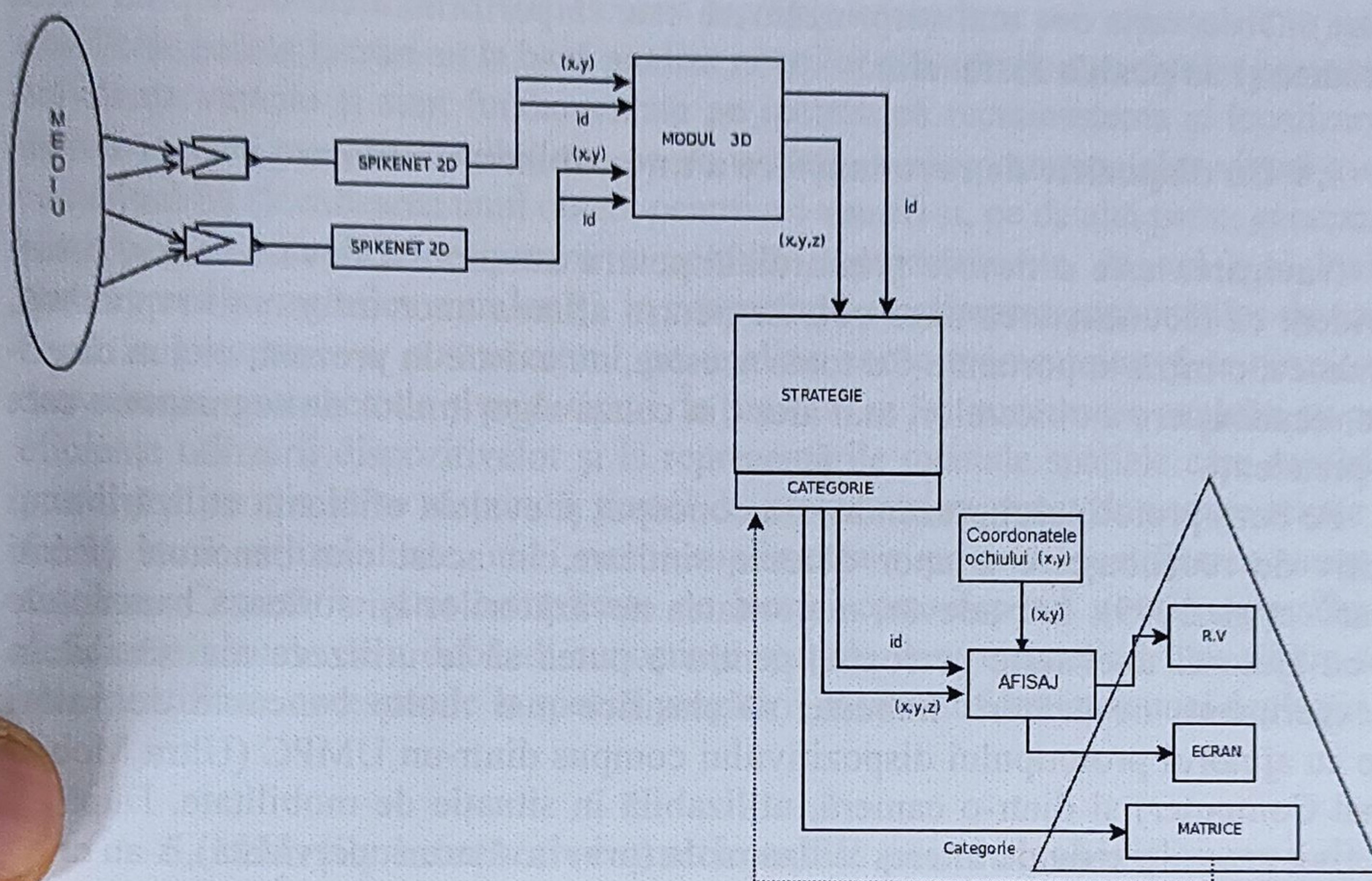


Fig. 12. Arhitectura schematică a dispozitivului experimental.

S-a dezvoltat o arhitectură care permite modificarea modelelor de fosfene afișate în casă: acestea depind de locul de stimulare, de numărul și de poziția electrozilor, de frecvența de stimulare și de intercomunicările dintre electrozi (a se vedea Fig. 12). Toți acești parametri se bazează pe analiza literaturii cu privire la stimularea electrică a cortexului vizual la primate.

Figura 12 prezintă arhitectura schematică a dispozitivului experimental. Imaginile provenite de la camere sunt modificate de un soft de recunoaștere a obiectelor. Notăția Id corespunde identității obiectului a cărui poziție în 2D, apoi în 3D este comunicată modulului „strategii/categorii” care va transforma această informație în fosfene. Acest modul permite, în special, modificarea modelului de matrice a electrodului utilizat (mărime și configurație), precum și poziția sa în sistemul vizual. Modelele de fosfene evocate depind, în mod evident, de tipul și de poziția matricei.

Doisprezece subiecți (6 bărbați și 6 femei, cu vârsta medie de 24 de ani) au participat la un experiment de recunoaștere și de localizare a unor obiecte (Guivarch⁸⁹ 2010). Sarcina consta în apucarea unui obiect plasat pe o masă printre alte obiecte prin intermediul unei vederi protetice prin localizare simulată (a se vedea Fig. 11, stânga). Subiecții trebuiau să realizeze această sarcină utilizând matrice de mărimi diferite (9 sau 90) și localizări de implantare a matricelor diferite (centrală sau excentrată).

⁸⁹ Guivarch, Valérian, *Modélisation d'une neuroprothèse de reconnaissance et de localisation d'objet dans l'espace*, Master IT-ICSC: University of Toulouse 3, 2010.

Rezultatele arată că rata de succes (82% în medie) și duratele până la apucare (18 s) sunt excelente oricare ar fi obiectul căutat (deși unele sunt mai bine detectate de sistem). Observația cea mai surprinzătoare este că performanțele nu depind deloc de cantitatea și de poziția fosfenelor.

4.4 Un dispozitiv de recunoaștere a unor obiecte similare

Navigarea este o nevoie primordială pentru autonomia nevăzătorilor, totuși este evident că recunoașterea unor obiecte pentru aflarea unor informații reprezintă, de asemenea, o miză importantă. Cu toate acestea, nu există, în prezent, niciun dispozitiv de recunoaștere a obiectelor, mai ales din cauza algoritmilor de segmentare care rămân prea lenți.

Pe baza prototipului prezentat, s-a conceput și evaluat eficiența utilizării unui dispozitiv de recunoaștere a unor obiecte similare, în acest caz bancnote (Rémi Parlouar⁹⁰ et al. 2009). Într-adevăr, majoritatea nevăzătorilor își sortează bancnotele acasă, cu ajutorul unei terțe persoane, pentru a putea să le utilizeze mai târziu. În acest experiment, nevăzătorii trebuiau să clasifice mai multe bancnote de valori diferite cu ajutorul prototipului dispozitivului compus dintr-un UMPC (Ultra Mobile Personal Computer) și dintr-o cameră, utilizabilă în situație de mobilitate. După ce s-au definit anumite zone de interes la bancnote (prin învățare supervizată), s-au creat modele care permit recunoașterea bancnotelor în toate orientările. În final, experimentul consta în clasificarea bancnotelor în funcție de valoarea acestora.

În momentul interviului final, toți utilizatorii, cu excepția unuia, au găsit dispozitivul foarte util și utilizabil. Ei propun dezvoltarea acestui tip de aplicație pe telefonul mobil și extinderea sa la alte tipuri de obiecte cum ar fi CD-urile, ambalajele alimentare etc. Evident, o astfel de aplicație de recunoaștere a unor obiecte poate fi integrată cu ușurință într-un dispozitiv de asistare spațială. Această convergență de întrebări în cadrul aceleiași dispozitiv este un argument suplimentar pentru abordarea supleanței.

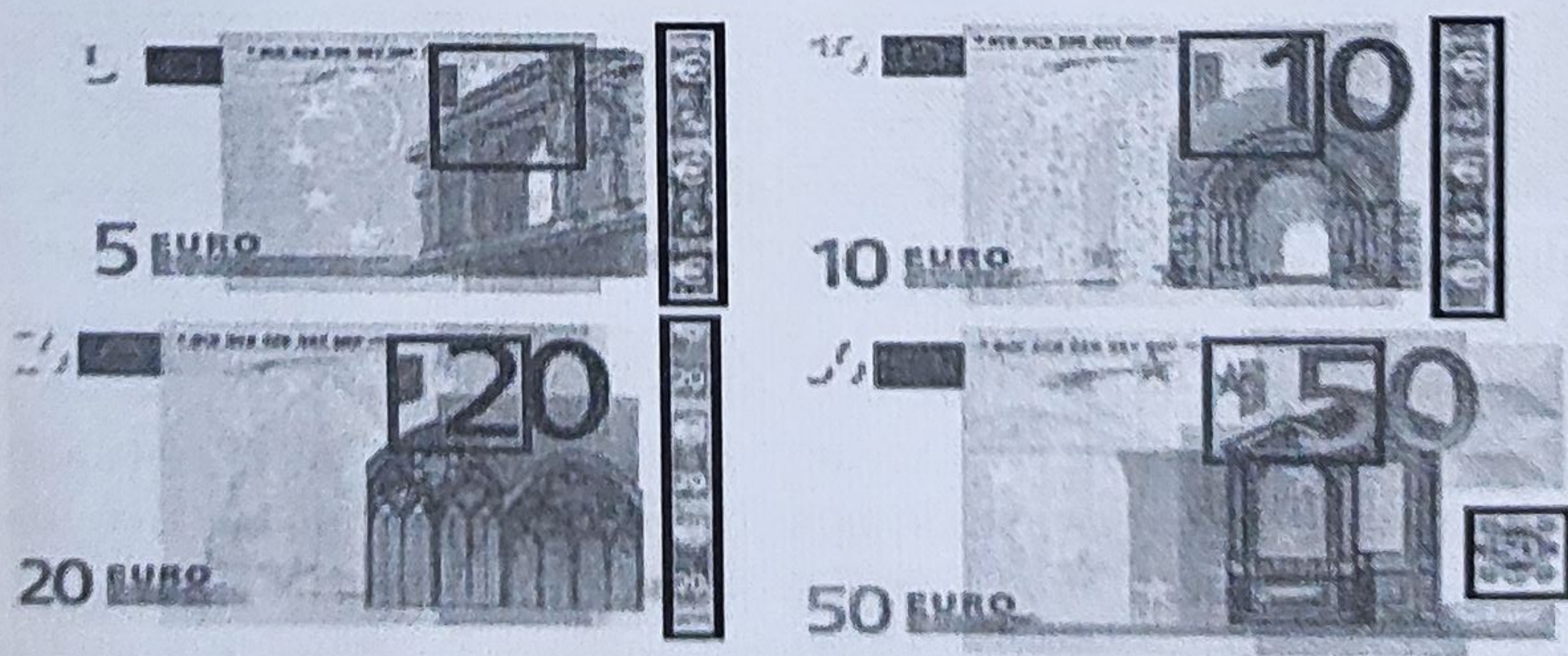


Fig. 13. Bancnotele de 5, 10, 20 și 50 de euro.
Cu roșu, zonele caracteristice învățate și recunoscute ale bancnotelor

⁹⁰ Parlouar, Rémi, Florian Dramas, Marc M. Macé & Christophe Jouffrais, *Assistive device for the blind based on object recognition: an application to identify currency bills*, Paper presented to the ASSETS'09, Pittsburgh, USA, 2009.

4.5 Concluzii privind supleanța spațială

În această secțiune, s-au descris lucrările întreprinse privind supleanța spațială in situ. Toate aceste lucrări au la bază analiza nevoilor și a comportamentelor persoanelor cu deficiențe vizuale și sunt fundamentate pe ipoteza că recunoașterea și localizarea unor indicii vizuale permite restabilirea, pe de o parte, a comportamentelor vizual-motorii în buclă închisă (localizarea unui obiect pentru a-l apuca) și, pe de altă parte, generarea unor hărți cognitive suficiente pentru a permite comportamente de cel mai înalt nivel (înțelegerea topografiei unui loc și, în consecință, ameliorarea capacităților de navigare). Până în prezent, s-a pus accentul pe analiza comportamentelor și a nevoilor și pe dezvoltarea unor metode și prototipuri. În prezent, se fac numeroase evaluări cu privire la eficiența utilizării dispozitivelor și la reprezentările mentale spațiale care rezultă. Acest prototip a permis să se răspundă, destul de ușor dintr-un punct de vedere tehnic și cu o foarte bună utilizabilitate, unei nevoi suplimentare privind identificarea unor obiecte similare. Cercetările sunt completate de lucrări care răspund nevoilor persoanelor cu deficiențe vizuale de a dobândi cunoștințe spațiale despre exterior. Una dintre sarcinile clar identificate corespunde în special pregătirii itinerarului înaintea unei deplasări.

5. Hărțile interactive multimodale pentru achiziționarea de cunoștințe spațiale

5.1 Introducere

Percepția prin simțul tactil este un mijloc prin care persoanele cu deficiențe vizuale pot achiziționa informații și, în special, informații spațiale. Datorită complementarității unei interacțiuni modale (audio-tactile) pentru persoanele cu deficiențe vizuale, s-a realizat o colaborare cu Delphine Picard, profesor de psihologie la Universitatea le Mirail din Toulouse, în cadrul unui proiect privind percepția imaginilor prin simțul tactil la persoanele cu deficiențe vizuale.

Omul utilizează, probabil, de mai multe milenii, hărți și descrieri de itinerare pentru a naviga. În prezent, aceste resurse sunt puse la dispoziția indivizilor în formă digitală, ceea ce le-a făcut foarte populare, mai ales după inventarea Smartphone-urilor și a aplicațiilor (Google Maps, Mappy etc.) care oferă funcții noi pentru asistarea navigării. În esență, hărțile vizuale sunt inutilizabile pentru persoanele cu deficiențe vizuale. Totuși, ele sunt indispensabile în cadrul pregătirii itinerariilor, a fortiori pentru deplasarea unei persoane cu deficiențe vizuale. Într-adevăr, este deosebit de important pentru persoanele cu deficiențe vizuale să poată pregăti un traseu nou cu informații despre forma generală a acestui traseu, precum și despre punctele de reper de pe parcursul său. Această pregătire permite, în special, diminuarea temerilor și a anxietății datorate imersiunii într-un mediu necunoscut.

Simțul nostru tactil permite recuperarea informațiilor spațiale plecând de la o imagine imprimată în relief. Aceste „imagini tactile” (Eriksson⁹¹ 1998) pot avea o

⁹¹ Eriksson, Yvonne, *Tactile pictures: pictorial representations for the blind*, Acta Universitatis Gothoburgensis, 1998.

mare utilitate pentru persoanele cu deficiențe vizuale deoarece permit obținerea unor informații cu privire la pozițiile, mărimea și formele obiectelor (Kennedy⁹² 1993). Un rezultat recurent, dar nesistematic arată că orbii tardivi au rezultate superioare celor ale orbilor din naștere (Heller⁹³ 1989).

Primul obiectiv al acestui proiect era de a înțelege capacitățile perceptive și cognitive ale persoanelor cu deficiențe vizuale în momentul explorării unor imagini fără semnificație, apoi cu semnificație. După această primă etapă, care permite o bună înțelegere a mecanismelor percepției simțului tactil la nevăzători și la văzători, s-a pus accentul pe explorarea unor hărți tactile și pe conceperea unui dispozitiv interactiv multimodal care să permită persoanelor cu deficiențe vizuale să citească hărți geografice.

5.2 Percepția imaginilor prin simțul tactil

„Imaginile tactile” pot fi create, de exemplu, plecând de la materiale sensibile la căldură (hârtie Swell) sau sensibile la presiune (foi Mylar). În ambele cazuri, urmele grafice apar în relief (1 mm) și, prin urmare, sunt perceptibile cu degetele mâinii. Problema accesibilității imaginilor în sistemul tactil face obiectul unei dezbateri recente. Într-adevăr, specialiștii în domeniul pipăitului au susținut mult timp că simțul pipăitului este potrivit, în principal, pentru înțelegerea obiectelor tridimensionale și a calităților de suprafață ale acestor obiecte (a se vedea, de exemplu, Lederman & Roberta L. Klatzky⁹⁴ 1987; Révész⁹⁵ 1950).

Alte studii au arătat totuși că persoane atinse de cecitate vizuală pot, într-o anumită măsură, să înțeleagă și să producă desene tactile ale unor obiecte comune (Heller⁹⁶ 1989; Kennedy⁹⁷ 1993). Capacitățile de identificare ale desenelor tactile au fost testate și pe persoane văzătoare, cu un succes temperat (Heller⁹⁸ et al. 2005). Astfel, imaginea ar putea să capete un sens prin pipăit, depășind domeniul de percepție clasic vizual al acestui obiect bidimensional. Doar că mecanismele legate de percepția și de identificarea desenelor cu trăsături în relief prin simțul pipăitului sunt, în prezent, neînțelese bine și au nevoie de investigații noi.

În studiile realizate, s-a examinat rolul experienței vizuale și al reprezentării vizuale a imaginilor în analiza pattern-urilor tactile 2D simple (fără convenții de desen, precum perspectiva sau ocluzia). S-a comparat capacitatea orbilor precoci și tardivi și a văzătorilor cu ochii acoperiți într-o sarcină de recunoaștere a pattern-urilor 2D. S-a analizat, în special, dacă recunoașterea acestor pattern-uri depinde de tipul de strategie de memorare utilizat (de exemplu, spațial, vizual-spațial, verbal sau

⁹² Kennedy, John M., *Drawing and the blind: pictures to touch*, Yale University Press, 1993.

⁹³ Heller, Morton A., *Picture and pattern perception in the sighted and the blind: the advantage of the late blind*, *Perception* 18,379-89, 1989.

⁹⁴ Lederman, Susan J. & Roberta L. Klatzky, *Hand movements: a window into haptic object recognition*, *Cogn Psychol* 19,342-68, 1987.

⁹⁵ Révész, Géza, *Psychology and art of the blind* Oxford, England Longmans, Green, 1950.

⁹⁶ Heller, Morton A., *Picture and pattern perception in the sighted and the blind: the advantage of the late blind*, *Perception* 18,379-89, 1989.

⁹⁷ Kennedy, John M., *Drawing and the blind: pictures to touch*, Yale University Press, 1993.

⁹⁸ Heller, Morton A., Melissa McCarthy & Ashley Clark, *Pattern perception and pictures for the blind*, *Psicologica* 26,161-71, 2005.

kinestezic). S-a arătat că, în ciuda unor diferențe semnificative între grupuri în tipul de strategie utilizat, nu există nicio diferență de performanță. Totuși, rata de recunoaștere este corelată pozitiv cu capacitățile de reprezentare vizual-spațială a imaginilor ale subiecților văzători. Aceste rezultate permit să se concluzioneze că dificultățile anumitor nevăzători de a recunoaște imagini tactile nu se datorează unor dificultăți bazice în ceea ce privește analiza unor pattern-uri tactile 2D simple. Interesant este faptul că percepția imaginilor tactile 2D ar putea fi ameliorată ameliorând capacitățile de reprezentare vizual-spațială a imaginilor ale nevăzătorilor.

5.2.1 Recunoașterea prin simțul tactil a unor imagini fără semnificație

Mai multe lucrări arată că percepția imaginilor tactile de către persoanele cu deficiențe vizuale variază în funcție de natura imaginii în relief și de exigențele sarcinii. În timp ce formele geometrice simple (ca un dreptunghi) sunt recunoscute cu ușurință (Heller⁹⁹ 1989), identificarea se dovedește a fi mai dificilă atunci când desenul reprezintă structura tridimensională a unui obiect urmând convențiile vizuale (Lederman¹⁰⁰ et al. 1990). O constatare recurentă este că, în cazul adulților, orbii tardivi și văzătorii îi depășesc deseori pe orbii precoci. Aceste rezultate se înscriu în discuția actuală în cadrul căreia se confruntă două teorii cu privire la rolul experienței vizuale și al reprezentării vizuale a imaginilor în percepția unor imagini tactile. Teoria propusă de Lederman și colaboratorii (Lederman¹⁰¹ et al. 1990) sugerează că identificarea imaginilor tactile este mediată prin reprezentarea vizuală a imaginilor. În schimb, Kennedy (Kennedy¹⁰² 1993) susține ideea că identificarea de imagini tactile este posibilă fără vreo experiență vizuală și fără o reprezentare vizuală a imaginilor, deși destul de dificil; prin urmare, ar putea fi amodală.

S-a abordat această discuție punându-se problema rolului experienței vizuale și al reprezentării vizuale a imaginilor în analiza unor imagini tactile 2D (Delphine Picard¹⁰³ et al. 2010). Formele 2D fără semnificație sunt utile pentru a controla faptul că eventualele dificultăți de recunoaștere nu sunt legate de dificultăți de interpretare.

Au fost comparate performanțele unor orbi precoci (OP), tardivi (OT) și a unor văzători cu ochii acoperiți (V) în timpul unui test de recunoaștere a unor imagini 2D fără semnificație (Fig. 14). S-a analizat și dacă există vreo corelație între recunoașterea acestor imagini tactile 2D și tipul de strategie mnezică utilizat de participanții OP, OT și V. Nu s-a remarcat nicio diferență semnificativă între grupuri în ceea ce privește recunoașterea și, aceasta, în ciuda utilizării unor strategii diferite.

⁹⁹ Heller, Morton A., *Picture and pattern perception in the sighted and the blind: the advantage of the late blind*, Perception 18.379-89, 1989.

¹⁰⁰ Lederman, Susan J., Roberta L. Klatzky, Cynthia Chataway & Craig D. Summers, *Visual mediation and the haptic recognition of two-dimensional pictures of common objects*, Percept Psychophys 47.54-64, 1990.

¹⁰¹ Lederman, Susan J., Roberta L. Klatzky, Cynthia Chataway & Craig D. Summers, *Visual mediation and the haptic recognition of two-dimensional pictures of common objects*, Percept Psychophys 47.54-64, 1990.

¹⁰² Kennedy, John M., *Drawing and the blind: pictures to touch*, Yale University Press, 1993.

¹⁰³ Picard, Delphine, Samuel Lebaz, Christophe Jouffrais & Catherine Monnier, *Haptic recognition of two-dimensional raised-line patterns by early-blind, late-blind, and blindfolded sighted adults*, Perception 39.224-35, 2010.

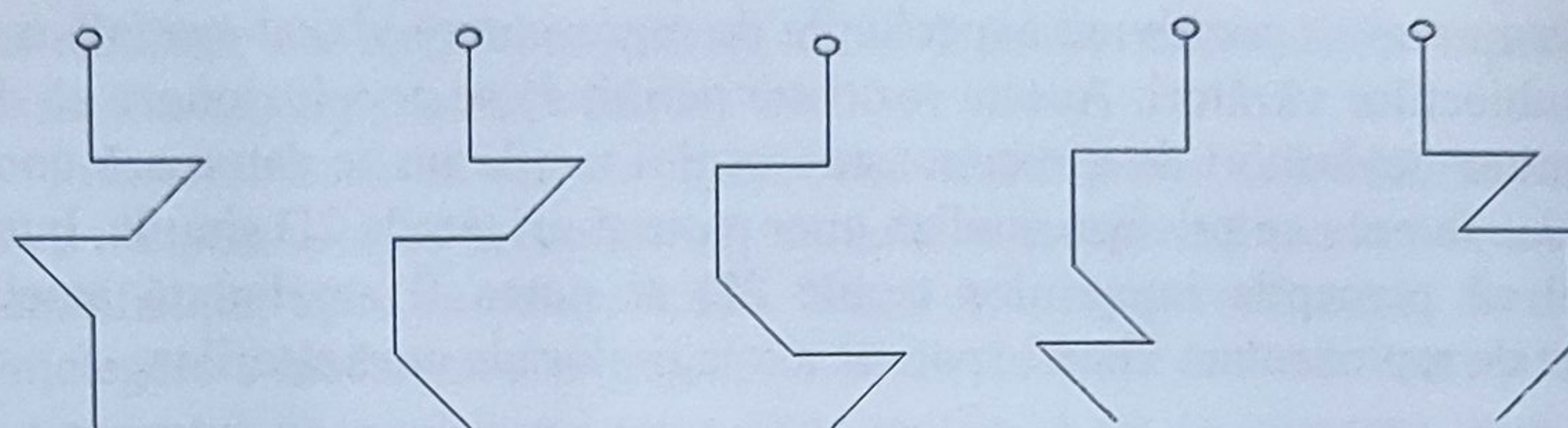


Fig. 14. Exemple de imagini fără semnificație (Delphine Picard et al, 2010).

5.2.2 Recunoașterea prin simțul tactil a unor imagini cu semnificație

Sistemul haptic permite identificarea unor obiecte comune. Această identificare este foarte performantă (aproape 100%) și are la bază mecanisme de explorare care permit o identificare în 1 până la 2 secunde (Roberta L. Klatzky¹⁰⁴ et al. 1985). Dimpotrivă, sistemul haptic este mult mai puțin eficient în ceea ce privește identificarea unor desene cu trăsături în relief ale unor obiecte comune. Procesul este mult mai lung deoarece este de ordinul a 90 de secunde, cu o performanță care nu depășește 33% (Lederman¹⁰⁵ et al. 1990). În modelul de identificare prin simțul tactil propus de Lederman¹⁰⁶ et al. (1990), recunoașterea unor desene cu trăsături în relief implică patru etape, dintre care o etapă de mediere vizuală.

5.2.2.1 Obiecte comune

S-a elaborat un studiu (Samuel Lebaz¹⁰⁷ et al. 2011) care confirmă că reprezentarea vizuală a imaginilor este implicată în recunoașterea prin simțul tactil a unor desene cu trăsături în relief ale unor obiecte comune. Pentru a măsura capacitățile subiecților de reprezentare a imaginilor, s-a utilizat versiunea revizuită a *Minnesota Paper Form Board* (MPFB, Likert & Quasha¹⁰⁸ 1941). Apoi, s-au corelat performanțele de identificare prin simțul tactil cu capacitățile de reprezentare vizual-spațială a imaginilor. Rezultatele evidențiază faptul că cei cu o reprezentare vizual-spațială a imaginilor mai puternică au performanțe mai bune de recunoaștere decât cei cu o reprezentare vizual-spațială a imaginilor mai slabă. Însă, timpii lor de răs-

¹⁰⁴ Klatzky, Roberta L., Susan J. Lederman & Victoria A. Metzger, *Identifying objects by touch: an "expert system"*, *Percept Psychophys* 37, 1985.

¹⁰⁵ Lederman, Susan J., Roberta L. Klatzky, Cynthia Chataway & Craig D. Summers, *Visual mediation and the haptic recognition of two-dimensional pictures of common objects*, *Percept Psychophys* 47.54-64, 1990.

¹⁰⁶ Lederman, Susan J., Roberta L. Klatzky, Cynthia Chataway & Craig D. Summers, *Visual mediation and the haptic recognition of two-dimensional pictures of common objects*, *Percept Psychophys* 47.54-64, 1990.

¹⁰⁷ Lebaz, Samuel, Christophe Jouffrais & Delphine Picard, *Haptic identification of raised-line drawings: high visuospatial imagers outperform low visuospatial imagers*, *Psychol Res.* 1-9, 2011.

¹⁰⁸ Likert, Rensis & William H Quasha, *Minnesota Paper Form Board Test*, The Psychological Corporation, 1941.

puns sunt echivalenți. În acest studiu, s-au utilizat două tipuri de desene cu trăsături în relief care diferă prin prezența sau nu a unor indicii 3D, ca unele distorsiuni de proiectare (Fig. 15). Conform rezultatelor lui Lederman¹⁰⁹ et al. (1990), subiecții erau mai buni și mai rapizi în identificarea desenelor 2D decât 3D. Surprinzător, capacitățile de reprezentare vizual-spațială a imaginilor au avut un efect asupra recunoașterii 2D prin simț tactil, dar nu și asupra identificării desenelor 3D.

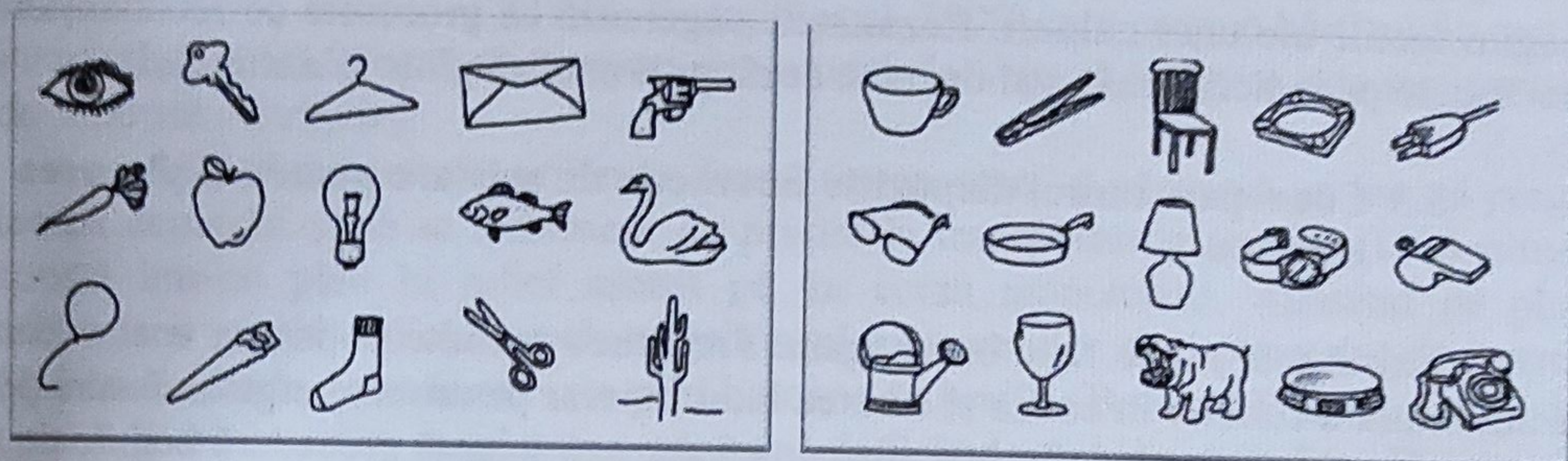


Fig. 15. Desene 2D (stânga) și 3D (dreapta) utilizate în studiul lui Samuel Lebaz & al., 2010

5.2.2.2 Clasificarea emoțiilor pe chipuri (pe figură)

În final, o problemă deosebit de interesantă a fost aceea de a afla dacă adulții nevăzători pot înțelege expresii faciale emoționale explorând imagini cu trăsături în relief. Prin urmare, s-a elaborat un ultim studiu în care adulți văzători și orbi trebuiau să clasifice imagini reprezentând chipuri ce exprimau diferite emoții. În timp ce rata de succes nu varia semnificativ între cele două grupuri, adulții orbi erau mai rapizi. S-a arătat și că există corelații puternice între clasificările realizate de orbi și de văzători. Acest lucru confirmă faptul că anumite imagini tactile pot fi înțelese fără a recurge la reprezentări ale imaginilor sau la experiența vizuală (Heller¹¹⁰ 1989; Kennedy¹¹¹ 1993; Delphine Picard¹¹² et al. 2010).

5.2.3 Concluzii cu privire la percepția imaginilor prin simțul tactil

Lucrările precedente au arătat că nevăzătorii au performanțe echivalente cu cele ale subiecților văzători pentru a percepe forme tactile simple, fără semnificație. Prin urmare, se pare că nu există, la orbi, un deficit deosebit în mecanismele de bază ale percepției prin simțul tactil. Atunci, eventualele diferențe de performanță în ceea ce privește recunoașterea observate între orbi și văzători ar putea proveni din

¹⁰⁹ Lederman, Susan J., Roberta L. Klatzky, Cynthia Chataway & Craig D. Summers, *Visual mediation and the haptic recognition of two-dimensional pictures of common objects*, *Percept Psychophys* 47.54-64, 1990.

¹¹⁰ Heller, Morton A., *Picture and pattern perception in the sighted and the blind: the advantage of the late blind*, *Perception* 18.379-89, 1989.

¹¹¹ Kennedy, John M., *Drawing and the blind: pictures to touch*, Yale University Press, 1993.

¹¹² Picard, Delphine, Samuel Lebaz, Christophe Jouffrais & Catherine Monnier, *Haptic recognition of two-dimensional raised-line patterns by early-blind, late-blind, and blindfolded sighted adults*, *Perception* 39.224-35, 2010.

conținutul semantic atașat imaginilor sau din codurile picturale care le compun. Studiul percepției de imagini cu semnificație la văzători evidențiază faptul că capacitățile de reprezentare a imaginilor sunt corelate cu performanța în ceea ce privește recunoașterea imaginilor 2D, însă nu și a imaginilor 3D. Este probabil că percepția imaginilor 3D este dependentă de mecanisme de integrare suplimentare.

În final, s-a arătat că nevăzătorii sunt capabili să recunoască emoții pe imagini tactile ale unor chipuri. Rezultatele sugerează că procesele de recunoaștere sunt bazate pe indicii locale mai degrabă decât pe o analiză globală a chipului.

5.3 Conceperea unui dispozitiv interactiv de asistare pentru explorarea hărților

Există mai multe faze în navigare. Pregătirea permite obținerea unei vederi de ansamblu a zonelor în cauză și crearea unui itinerar pentru a ajunge la destinație. De obicei, persoanele cu deficiențe vizuale efectuează această procedură înainte de a se deplasa, într-un mediu sigur. Hărțile tactile sunt utilizate în educația persoanelor cu deficiențe vizuale de mult timp pentru a transmite idei care au o componentă spațială. Ele au fost utilizate ca ajutor pentru orientare (Golledge¹¹³ 1991) și ca ajutor pentru învățarea mobilității (Ungar¹¹⁴ et al. 1993; Ungar¹¹⁵ et al. 1995). Hărțile tactile permit extinderea înțelegerii unui mediu dincolo de ceea ce permite experiența directă. Cu toate acestea, o hartă fără etichete și legende este inutilă, ceea ce pune probleme în conceperea și utilizarea hărților tactile. Într-adevăr, etichetarea în Braille nu este flexibilă și devine repede supărătoare (Tatham & Dodds¹¹⁶ 1988). Iar utilizarea de etichete într-o legendă separată reduce rapiditatea percepției și introduce probleme de interpretare (Hinton¹¹⁷ 1993).

În prezent, există numeroase instrumente disponibile pe internet (de exemplu Google Maps) care permit consultarea de hărți și selectarea de itinerare. În plus față de hărțile vizuale, aceste instrumente dau indicații textuale de ghidare sub forma unor foi de drum (itinerar descris într-un reper egocentric) care pot fi accesibile datorită unui cititor de ecran. Totuși, după cum s-a precizat anterior, aceste instrumente au la bază sisteme informaționale geografice concepute pentru vehicule. Acestea nu permit generarea de itinerare pentru pietoni, cu atât mai puțin pentru pietoni cu deficiențe vizuale. În toate cazurile, aceste instrumente nu permit dobândirea unei reprezentări alocentrice a mediului care este suportul necesar aplicarea unor strategii cognitive. O soluție ar fi să se conceapă o hartă digitală multimodală (nevizuală) care să includă itinerariul dorit.

¹¹³ Golledge, Reginald G., *Tactual strip maps as navigational aids*, Journal of Visual Impairment & Blindness 85.296-301, 1991.

¹¹⁴ Ungar, Simon, Mark Blades & Christopher Spencer, *The role of tactile maps in mobility training*, British Journal of Visual Impairment 11.59, 1993.

¹¹⁵ Ungar, Simon, Mark Blades & Christopher Spencer, *Visually impaired children's strategies for memorising a map*, British Journal of Visual Impairment 13.27-32, 1995.

¹¹⁶ Tatham, Andrew P. & Allan G. Dodds, *Proceedings of the Second International Symposium on Maps and Graphics for Visually Handicapped People*, Paper presented to the Second International Symposium on Maps and Graphics for Visually Impaired People Nottingham, UK, 1988.

¹¹⁷ Hinton, Ron, *Tactile and audio-tactile images as vehicles for learning*, Paper presented to the Non-visual human-computer interactions: prospects for the visually handicapped, Paris, 1993.

Heuten¹¹⁸ et al (2006) au propus o interfață de sonificare 3D pentru a explora hărțile orașelor. Datorită unei navigări cu ajutorul tastaturii, fiecare tip de obiect geografic și fiecare tip de reper sunt asociate unui sunet și, prin urmare, pot fi identificate.

Potrivit lui Golledge (Golledge¹¹⁹ et al. 2005) adăugarea de informații sonore hărților în relief le face mai ușor de înțeles. Există câteva proiecte¹²⁰ (Gallagher & Frasch¹²¹ 1998; Jacobson¹²² 1998; Wang¹²³ et al. 2009) care propun hărți audio-tactile interactive ce au la bază un support monotouch (o singură atingere este recunoscută de suportul inetractiv).

ABApans permite utilizatorilor să se repereze, să găsească un loc, să pregătească deplasări și să se informeze cu privire la transporturile publice. Dispozitivul constă într-un plan în relief așezat pe un ecran monotouch. Apăsând pe plan, comentarii sonore indică numele străzilor, al intersecțiilor și al punctelor de interes. Mai multe moduri pot fi activate prin butoane în braille. Butonul P activează modul „plan”, care permite explorarea planului indicând numele străzilor. Butonul O activează modul orientare, care dă informații cu privire la distanțe și direcții în raport cu un punct de referință. Butonul I permite parcurgerea itinerariilor. În sfârșit, butonul T dă acces la informații privind transporturile publice.

Cu toate acestea, tehnologia monotouch are importante limitări în ceea ce privește interacțiunea și prezentarea informației. O interacțiune pe un support multitouch ar permite noi posibilități, cum ar fi adăugarea de interacțiuni de tip zoom sau scrolling, sau de interacțiuni suplimentare care permit calcularea unor itinerare bis, distanțe, explorarea și oferirea de informații cu privire la locuri (de exemplu restaurante sau muzee) etc. Se pare că nu există niciun proiect de hartă interactivă multitouch, care să permită, în special, conceperea și evaluarea a noi metode de interacțiune pentru persoanele cu deficiențe vizuale.

¹¹⁸ Heuten, Wilko, Daniel Wichmann & Susanne Boll, *Interactive 3D sonification for the exploration of city maps*, Paper presented to the NordiCHI 2006, 2006.

¹¹⁹ Golledge, Reginald G., Matthew Rice & Daniel R. Jacobson, *A Commentary on the Use of Touch for Accessing On-Screen Spatial Representations: The Process of Experiencing Haptic Maps and Graphics*, *The Professional Geographer* 57.339-49, 2005.

¹²⁰ A se vedea, de exemplu :

- ABAPans : planuri ale orașelor pentru persoane oarbe și cu deficiențe de vedere, disponibile la adresa <http://abapans.eig.ch/>

- The Map Room : http://www.maproomblog.com/207/03/maps_for_the-visually_impaired.php

¹²¹ Gallagher, Blaithin & Walter Frasch, *Tactile Acoustic Computer Interaction System: A new type of Graphic Access for the Blind*, Paper presented at the 3rd TIDE congress 1998, Helsinki, 1998.

¹²² Jacobson, Daniel R., *Navigating maps with little or no sight: An audio-tactile approach*, Paper presented at the Content Visualization and Intermedia Representations, University of Montreal, Montreal, 1998.

¹²³ Wang, Zheshen, Baoxin Li, Terri Hedgpeth & Teresa Haven, *Instant tactile-audio map: enabling access to digital maps for people with visual impairment*, Paper presented at the ASSETS'09, Pittsburgh, Pennsylvania, USA, 2009.

5.3.1 Concepția participativă

Hărțile pot fi utilizate în diverse moduri și cu diverse obiective. Pentru a fi în măsură să asiste persoanele cu deficiențe vizuale, este important să se înțeleagă utilizarea și proprietățile hărților orașului.

O sarcină importantă a utilizatorilor de planuri ale orașului este de a explora o regiune necunoscută și de a obține o idee cu privire la organizarea sa generală.

O a doua sarcină la rezolvarea căreia hărțile pot contribui este de a cunoaște împrejurimile unui amplasament specific. Utilizatorul vrea să știe care sunt obiectele geografice apropiate de acest loc și în ce direcție se află acestea.

O a treia sarcină are în vedere localizarea absolută într-un oraș a unor locuri importante, precum gările, parcurile, stațiile de transport public, centrele comerciale etc. Evident, hărțile au și rolul de a planifica un itinerar de la un punct de plecare către un punct de sosire. În acest cadru, este important să poți măsura distanțe relative sau absolute. În timpul unei deplasări, harta poate ajuta și la orientare.

O întrebare tipică la care utilizatorul dorește să răspundă este: unde sunt acum și în ce direcție merg? Prezenta lucrare analizează hărțile de oraș, care, de obicei, furnizează informații cu privire la caracteristicile geografice ale orașelor și cartierelor acestora, străzilor, stațiilor de transport public, clădirilor publice, piețelor, parcurilor, fluviilor și râurilor, monumentelor, hotelurilor și anumitor magazine.

Trebuie separate cele două cazuri de întrebuințări, cel al pregătirii unui itinerar în amonte și cel al ghidării în timpul deplasării, deoarece informațiile de care are nevoie persoana cu deficiențe vizuale diferă în cele două situații. Hărțile interactive multimodale plasează pregătirea itinerariului înaintea deplasării.

S-a dovedit că anumite informații sunt primordiale pentru o majoritate de persoane cu deficiențe vizuale care doresc să pregătească o deplasare. Aceste informații se referă, în special, la locurile care trebuie evitate în cursul deplasării, precum piețele și spațiile deschise care provoacă dezorientări, sensurile giratorii, zonele delimitate unde toate reperele au dispărut, trotuarele prea largi deoarece, în mod paradoxal, sunt deseori încărcate de terasele cafenelelor, trotuarele prea înguste care nu permit circulația cu bastonul etc. Dimpotrivă, alte spații trebuie privilegiate. Este vorba de intersecțiile simple, străzile calme, marile bulevarde etc. Astfel, aceste diverse zone au căpătat o importanță negativă sau pozitivă în funcție de valența lor și de preferințele diversilor utilizatori. O primă funcție a unei hărți interactive ar putea deci să se refere la asistarea planificării unui traseu, evitând zonele dezagreabile sau periculoase și incluzând zonele preferate.

În timpul planificării unui itinerar, utilizatorii caută o referință (o clădire cunoscută sau un punct cardinal), apoi o direcție generală în raport cu această referință. Informațiile care privesc numărul și numele străzilor întâlnite sunt utile.

Cu ocazia acestei analize, a fost abordat un aspect nou: în plus față de asistarea navigării, utilizatorilor le-ar plăcea să aibă o experiență agreabilă pregătind traseul și deplasându-se. Acest punct, în legătură cu experiența utilizatorului nu este luat prea des în considerare în conceperea dispozitivelor de asistare.

5.3.2 Prototip

Principiul care a fost reținut în ceea ce privește un prototip de hartă interactivă multimodală este cel al unei hărți pe hârtie cu trăsături în relief, așezată pe un ecran interactiv multitouch. Un astfel de prototip permite o mare flexibilitate în ceea ce privește configurările hărților, interacțiunile, testele și evaluările. Aceasta implică determinarea ecranelor interactive compatibile și producerea de hărți compatibile.

5.3.2.1 Hărți în relief

Imaginile în relief sunt obținute în mai multe moduri. Se utilizează, în special, două procedee: fie datorită unor materiale sensibile la căldură (hârtie Swell), fie prin presiune pe o folie de plastic deformabilă (hârtie Mylar). În ambele cazuri, urmele obținute prezintă un relief (1 mm) și devin perceptibile cu degetele mâinii.

Planurile în relief au o istorie lungă în domeniul geografiei (Golledge¹²⁴ et al. 2005). Reliefarea adaugă o dimensiune care aduce informații suplimentare hărților vizuale. Se utilizează planuri în relief pentru a reprezenta topografia. Evident, există și planuri în relief concepute special pentru persoanele cu deficiențe vizuale.

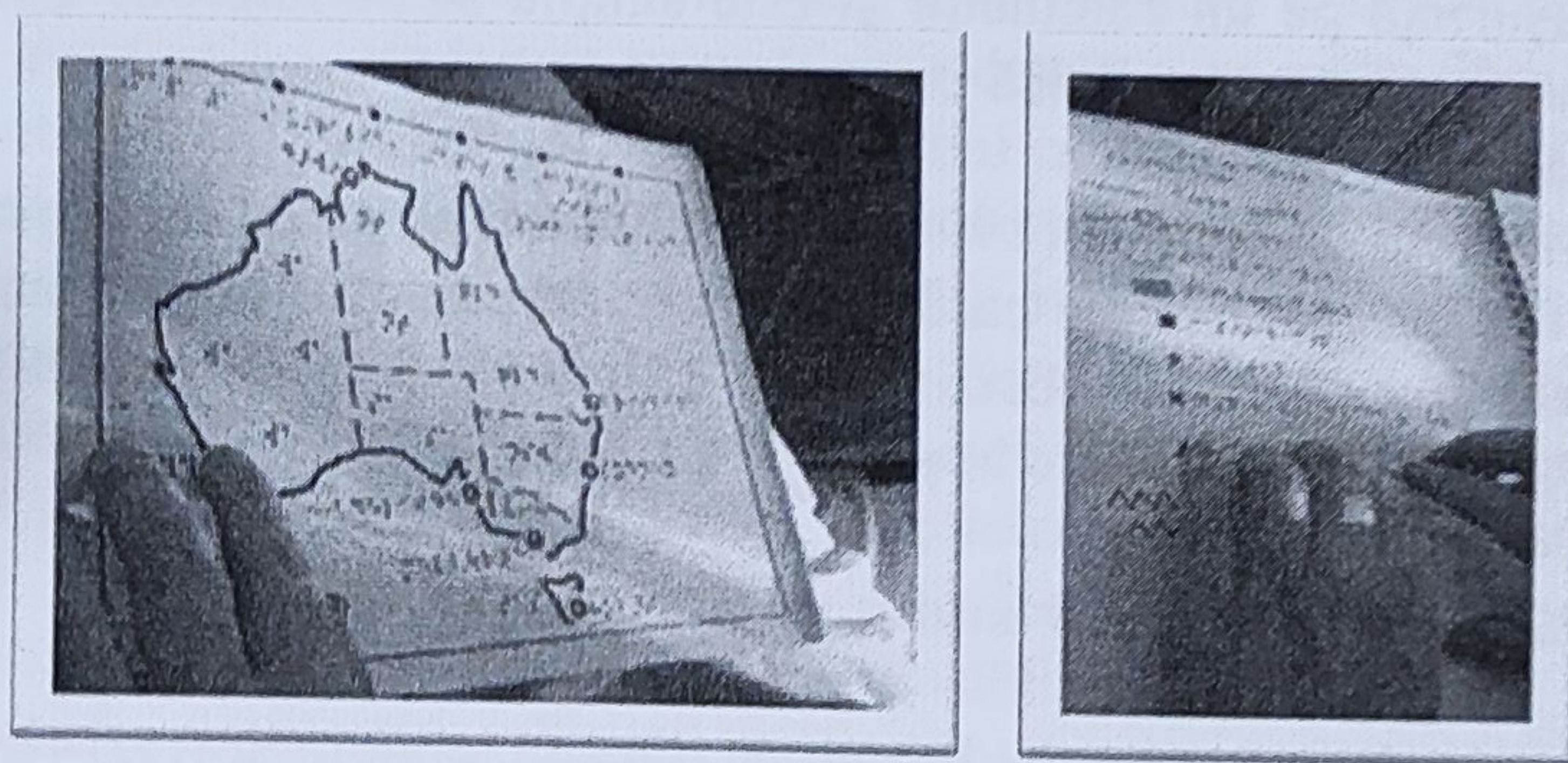


Fig. 11. Stânga: harta Australiei pe hârtie termoformată (se disting etichetele și legenda în Braille). Dreapta: caiet suplimentar, cu simbolurile utilizate, care însoțește hărțile.

Planurile conțin deseori etichete sub formă de abrevieri în Braille; în acest caz, ele necesită o legendă care explică abrevierile. Simbolurile utilizate diferă de la un plan la altul, prin urmare este necesar să se adauge o legendă care explică simbolurile utilizate (a se vedea Fig. 11). Utilizatorii comandă aceste produse la RNIB (Royal National Institute of Blind People) în Anglia sau la „The Princeton Braillelists” în Statele Unite ale Americii.

¹²⁴ Golledge, Reginald G., Matthew Rice & Daniel R. Jacobson, *A Commentary on the Use of Touch for Accessing On-Screen Spatial Representations: The Process of Experiencing Haptic Maps and Graphics*, *The Professional Geographer* 57.339-49, 2005.

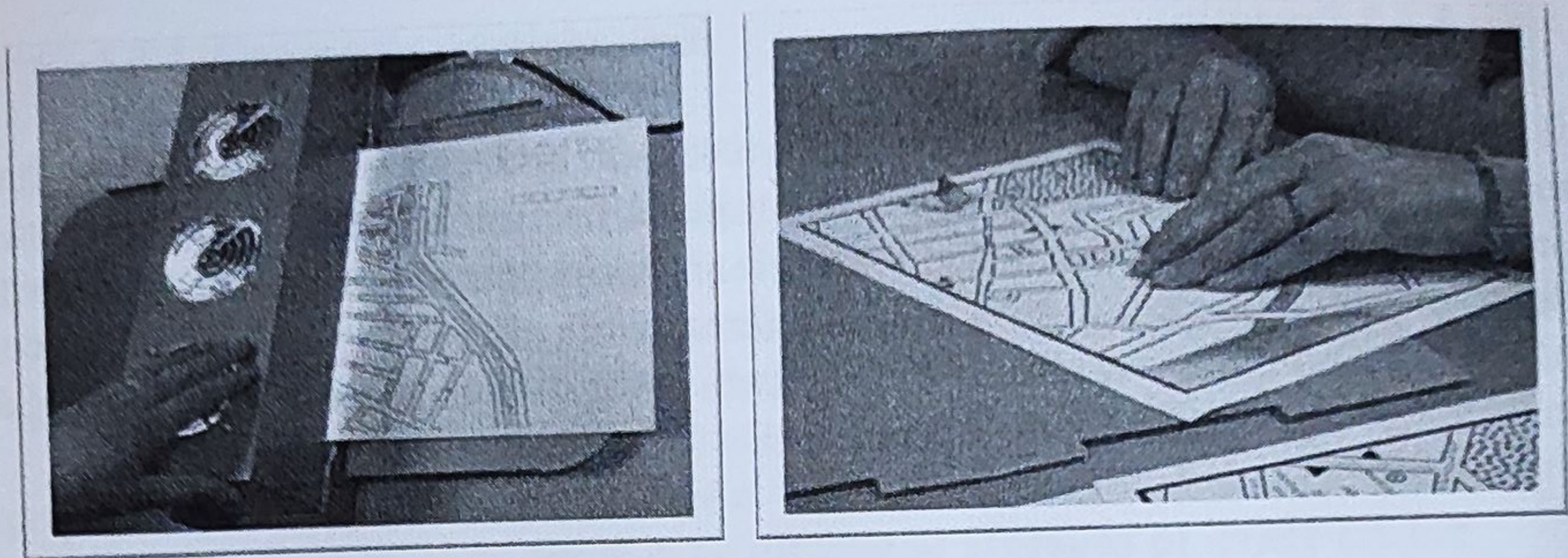


Fig. 17. Imaginea din stânga: *cuptor pentru producția de hărți termoformate*.
Imaginea din dreapta: *Harta unui cartier din Toulouse produsă la IRIT*.

Munca lui Anke Brock (Anke Brock¹²⁵ 2010) a permis instituirea, la IRIT, a întregului lanț de producție de hărți în relief. Metoda de producție pe care am ales-o este cea a hârtiei termoformate („swell paper”). Planul este imprimat pe o hârtie specială care conține în învelișul său microcapsule cu alcool. Când este trecută prin cuptor, capsulele de alcool se măresc și creează relieful (a se vedea Fig. 17). Această tehnică este, în primul rând, mai puțin costisitoare decât celelalte. Imaginile pot fi pregătite cu ușurință pe un calculator și este simplu să se imprime din nou aceeași imagine. De altfel, această tehnică este adaptată în mod special pentru utilizarea de hărți cu durată scurtă de viață.

În opinia lui Tatham¹²⁶ (1991) este preferabil să se limiteze numărul minim de mărci tactile, iar criteriile pe care el le propune pentru o bună concepție a hărții sunt: claritatea, relevanța și precizia; fiecare dintre aceste criterii se aplică pentru dimensiunile „obiectiv”, „așezare în pagină” și „producție”.

5.3.2.2 Ecran interactiv multitouch

Mai multe interfețe tactile bazate pe diferite tehnologii (rezistivă, capacitivă, FTIR, infraroșie, cu unde de suprafață etc.) au fost testate. În final, s-a optat pentru un ecran multitouch Stantum. Prototipul este compus din patru module conectate prin „middleware Ivy”: pilotul adaptat reacționează la evenimentele care sunt transmise hărții în format SVG. Elementul atins pe hartă declanșează un mesaj care activează ieșirea audio cu condiția ca aceasta să corespundă unei atingeri (ieșirea sunetului nu trebuie să fie declanșată în timpul explorărilor). Se utilizează o configurație modulară pe magistrală logică, ceea ce permite adăugarea cu ușurință a unor funcționalități (de exemplu, un afișaj Braille).

¹²⁵ Brock, Anke, *Cartes interactives multimodales pour piétons malvoyants*, Master thesis Univ. P. Sabatier, Toulouse, 2010.

¹²⁶ Tatham, Andrew F., *The design of tactile maps: theoretical and practical considerations*, Paper presented to the International Cartographic Association: mapping the nations, London, 1991.

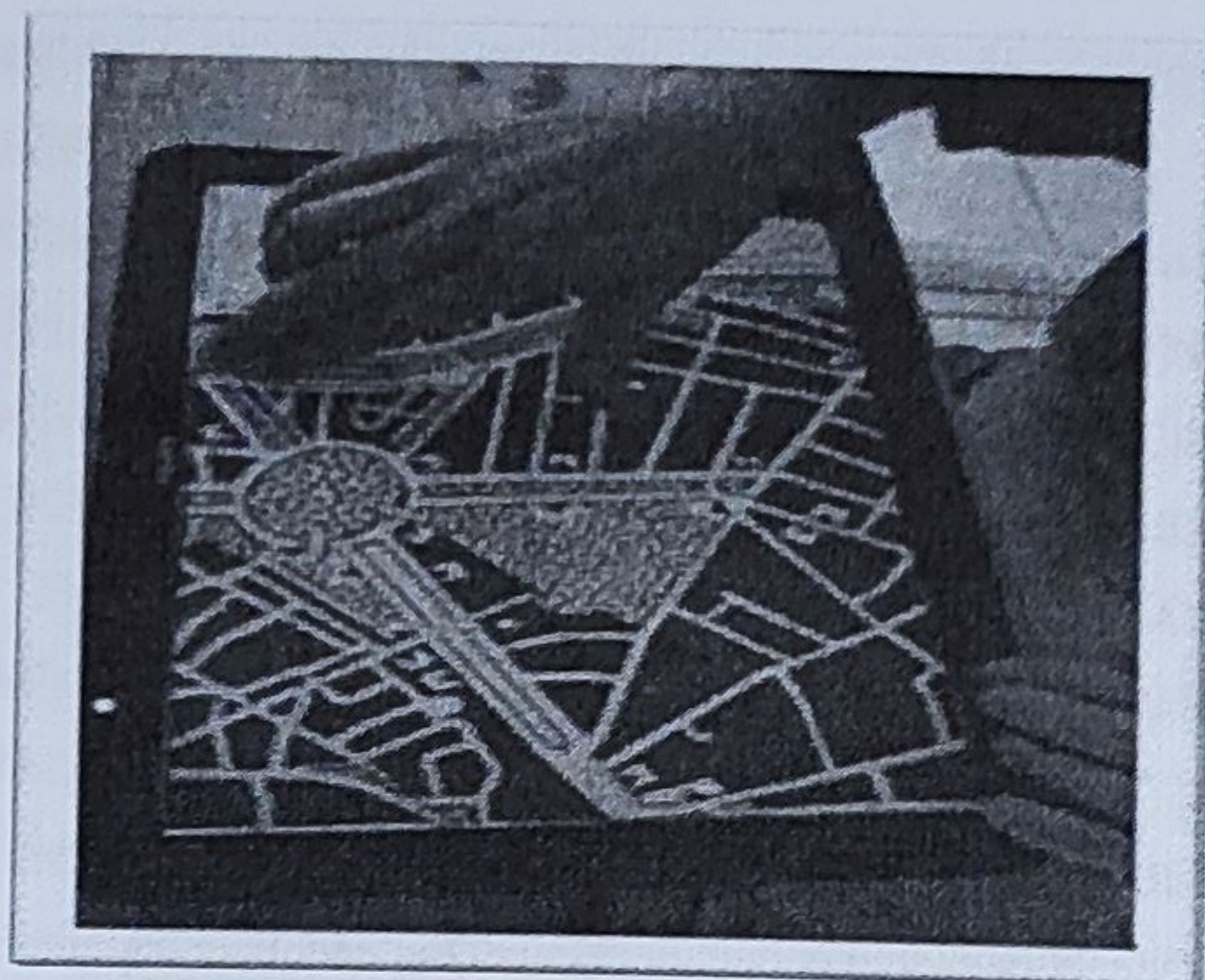


Fig. 18. *Prototip de hartă interactivă multimodală care include o hartă pe hârtie așezată pe un ecran multitouch. Elementele importante pot să transmită o informație sonoră și/sau în Braille. (imagine preluată din Anke Brock 2010 ITS)*

5.3.2.3 Captura mișcărilor degetelor

Un studiu recent pe adulți văzători (Wijntjes¹²⁷ et al. 2008) arată că sunt implicate multiple comportamente de explorare atunci când participanții trebuie să identifice imagini cu trăsăturile în relief ale unor obiecte comune (e.g., explorare dinamică cu două mâini, explorare cu un singur deget, a se vedea Fig. 19). Anumite comportamente permit, probabil, o explorare mai completă a imaginii decât altele. S-a dovedit că studiul strategiilor de explorare utilizate de văzători și de nevăzători ar putea avea consecințe importante asupra calității reprezentărilor mentale generate, precum și asupra concepției unui prototip de hartă interactivă multimodală. Prin urmare, s-a dezvoltat un modul de captură a mișcării și de recunoaștere a degetelor bazată pe Kinect. Captura traiectoriilor degetelor prin Kinect este cuplată la evenimentele de pe ecran. Ansamblul permite înregistrarea traiectoriilor fiecărui deget identificându-le. Prin urmare, va fi posibil să se analizeze strategiile de explorare a hărților tactile (Fig. 19).

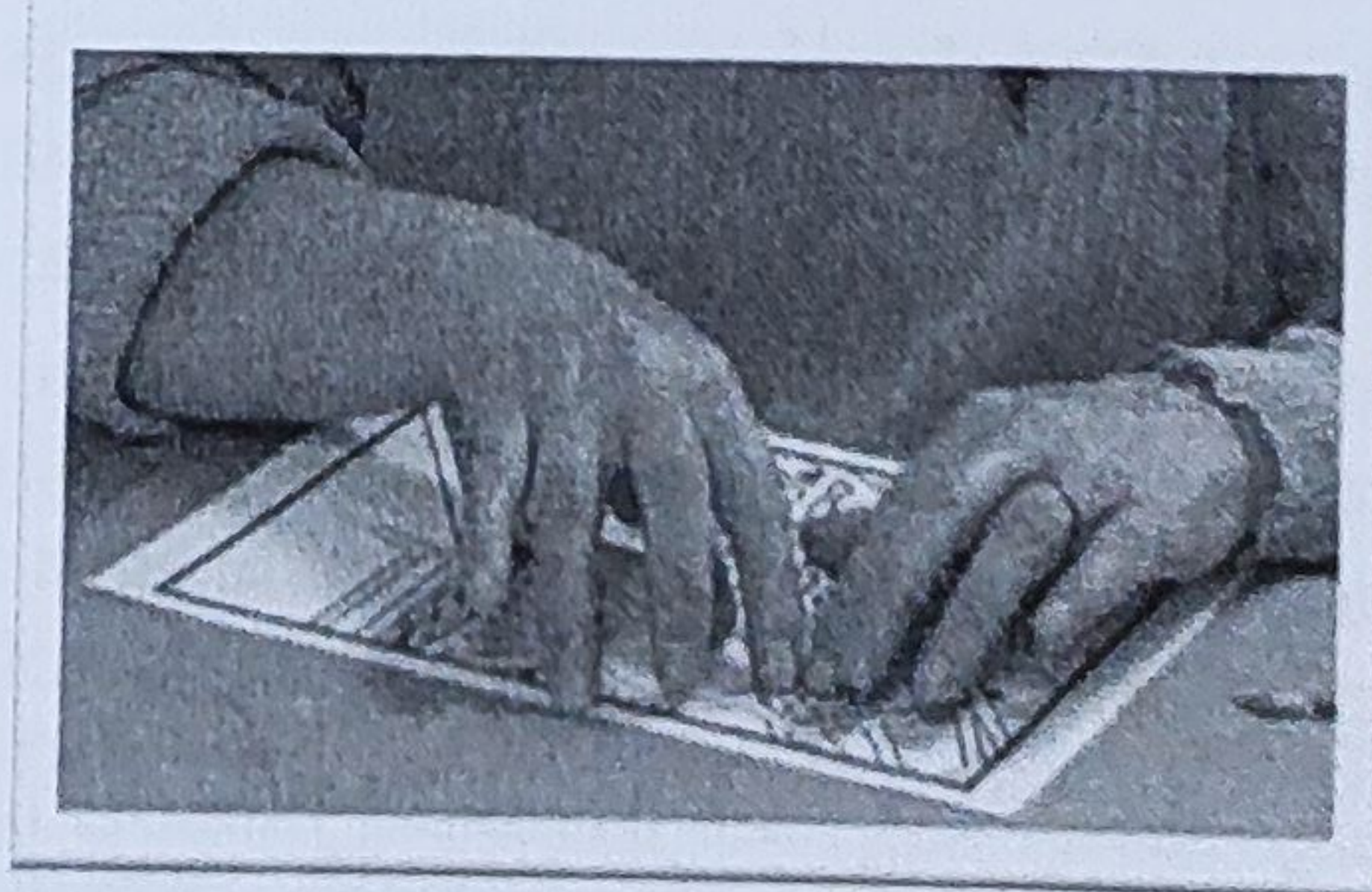


Fig. 19: *Exemple de strategii de explorare. Aici subiectul utilizează fie indexul unei singure mâini, fie ambele indexuri simultan.*

¹²⁷ Wijntjes, Maarten W., Thijs van Lienen, Ilse M. Verstijnen & Astrid M. Kappers, *The influence of picture size on recognition and exploratory behaviour in raised-line drawings*, Perception 37.602-14, 2008.

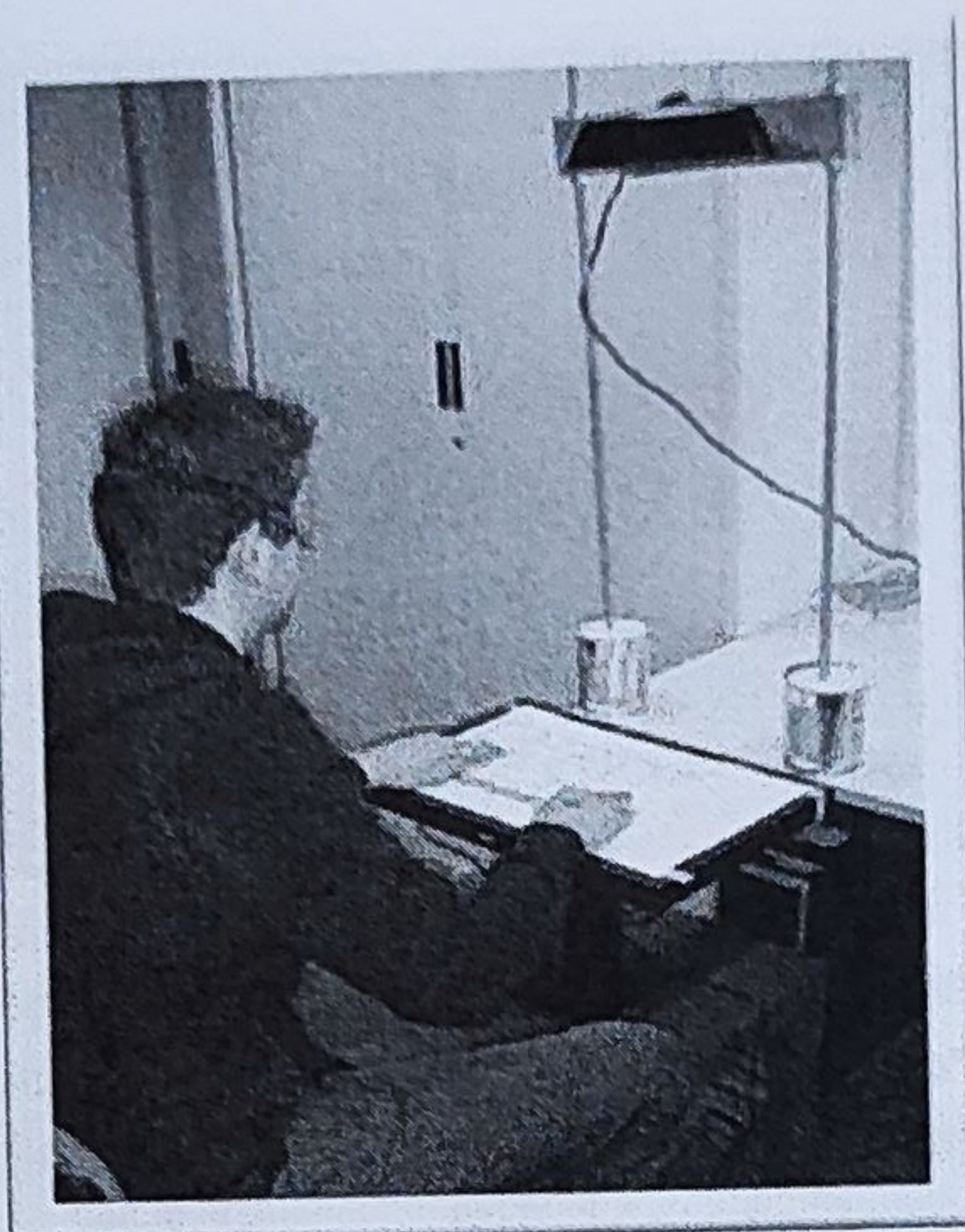


Fig. 20. *Dispozitiv experimental care permite studierea strategiilor de explorare a hărților tactile.*

Figura 20 prezintă un dispozitiv experimental care permite studierea strategiilor de explorare a hărților tactile. Dispozitivul include harta interactivă multimodală și un sistem de captură și de recunoaștere a degetelor.

Harta multimodală accelerează și ameliorează analiza informațiilor spațiale, pe de o parte, și ameliorează satisfacția utilizatorilor. 24 de subiecți nevăzători au fost testați în explorarea unor hărți pe hârtie și interactive, utilizând un ansamblu de întrebări referitoare la prezența unor repere, la itinerare și la configurațiile locurilor explorate, adresate imediat după explorare și la o săptămână după explorare. În colaborare cu Delphine Picard, s-au analizat rezultatele și s-a pus în evidență o absență a deficitelor perceptive la persoanele cu deficiențe vizuale și capacitatea acestora de a recunoaște imagini cu semnificație.

Concluzii generale

Primul rezultat important este colaborarea perenă a echipei ELIPSE cu Institutul Tinerilor Orbi din Toulouse, care permite punerea în aplicare a tuturor proiectelor de cercetare inițiate și, în consecință, înțelegerea cogniției spațiale a persoanelor cu deficiențe vizuale prin intermediul unor experiențe de psihologie experimentală. De asemenea, această colaborare permite conceperea unor dispozitive printr-o metodă de concepție participativă care include utilizatori finali și formatori de la centru. Această metodă, care a fost adaptată pentru a permite lucrul cu persoanele cu deficiențe vizuale (Anke Brock¹²⁸ et al. 2010b) se aplică atât evaluării unor tehnici de interacțiune specifice (localizare spațială prin sunete 3D, de

¹²⁸ Brock, Anke, Jean-Luc Vinot, Bernard Oriola, Slim Kammoun, Philippe Truillet & Christophe Jouffrais, *Méthodes et outils de conception participative avec des utilisateurs non-voyants*, Paper presented to the IHM'10, 2010b.

exemplu), cât și conceperii unor dispozitive complete (cum ar fi prototipul dispozitivului de asistare pentru navigare).

Lucrările realizate demonstrează că supleanța spațială prin localizare de obiecte datorită realității auditive mărite este o abordare funcțională. Într-adevăr, s-a arătat mai întâi că este posibil să se localizeze sunete reale în spațiul peripersonal în scopul de a poziționa mâna pe ținta sonoră (Florian Dramas¹²⁹ et al. 2008). Aceste lucrări sunt în curs de replicare într-un spațiu virtual (în colaborare cu LIMSI). Pe plan fundamental, ele vor duce la o mai bună definiție a indiciilor acustice care permit localizarea tridimensională a sunetelor. Pe planul IHM, vor permite conceperea unei metode de desemnare spațială în câmpul peripersonal. Rezultatele preliminare realizate cu sunete virtuale arată că este posibil să se utilizeze metafore speciale sau efecte pentru a ameliora localizarea în spațiul virtual apropiat. Această metodă de desemnare spațială va putea fi utilizată într-un ansamblu de sarcini primordiale pentru persoanele cu deficiențe vizuale. Se va putea recupera, în special, comportamentul vizual-motor de localizare și apucare a obiectelor¹³⁰. De asemenea, se vor putea indica eventuale obstacole pe traiectoria de apucare a obiectelor. În plus, s-a arătat deja că era posibil să se specifice direcțiile (azimut) prin localizarea unor ținte sonore în câmp îndepărtat (Roberta L. Klatzky¹³¹ et al. 2002). Această facultate ar putea fi utilizată pentru a recupera comportamentul senzorial-motor care constă în a se îndrepta către un reper îndepărtat, identificat în câmpul vizual. Acest lucru permite, de exemplu, să te îndrepti către o cutie poștală detectată în spațiul apropiat. Prototipul dispozitivului de asistare prin sinteză binaurală actual integrează aceste două funcții. El va fi evaluat într-un ansamblu de scenarii care au la bază localizarea de obiecte maimult sau mai puțin apropiate în scopul de a se apropia de ele sau de a le apuca. Interesant este faptul că dispozitivul facilitează convergența funcțiilor deoarece poate fi utilizat pentru a identifica obiecte similare (Rémi Parlouar¹³² et al. 2009). Această observație este foarte importantă deoarece, în ipoteza unei utilizări reale a dispozitivului, utilizatorii refuză, foarte logic, să ia cu ei mai multe dispozitive specifice pentru îndeplinirea unor sarcini diferite.

Acest studiu care se referă la ameliorarea poziționării și a ghidării pentru persoanele cu deficiențe vizuale scoate la iveală și alte două rezultate importante. Mai întâi, s-a conceput un sistem informațional geografic care permite integrarea nevoilor persoanelor cu deficiențe vizuale în ceea ce privește clacularea itinerariului

¹²⁹ Dramas, Florian, Brian F.G. Katz & Christophe Jouffrais, *Auditory-guided reaching movements in the peripersonal frontal space*, Paper presented at the The Journal of the Acoustical Society of America, 2008.

¹³⁰ A se vedea următorul film: <http://www.irit.fr/DFA-France/navig/ReachingToObjects.avi>. Acest film ilustrează o experiență preliminară care consta în apucarea telefonului mărit prin sunete virtuale. Experiența a fost filmată în condiții reale (fără montaj) cu numai trei modele vizuale de telefon și un sunet virtual 3D.

¹³¹ Klatzky, Roberta L., Yvonne Lippa, Jack M. Loomis & Reginald G. Golledge, *Learning directions of objects specified by vision, spatial audition, or auditory spatial language*, *Learning and Memory* 9.364-67, 2002.

¹³² Parlouar, Rémi, Florian Dramas, Marc M. Macé & Christophe Jouffrais, *Assistive device for the blind based on object recognition: an application to identify currency bills*, Paper presented to the ASSETS'09, Pittsburgh, USA, 2009.

și ghidarea (Slim Kammoun¹³³ et al. 2010). Acest system informațional geografic va fi utilizat în toate cercetările cu privire la asistarea electronică pentru pregătirea itinerariilor și navigarea persoanelor cu deficiențe vizuale. Al doilea rezultat important este că se pot utiliza repere geolocalizate pentru a ameliora poziționarea persoanelor cu deficiențe vizuale (Adrien Brilhault¹³⁴ et al. 2011). În prezent, s-au obținut rezultate suplimentare care arată că este posibil să se fixeze un traseu GPS datorită prezenței unui mic număr de repere pe parcurs.

Prototipul actual de dispozitiv integrează cele două funcții de „localizare a unor obiecte apropiate” și „ajutor pentru poziționare”, care sunt activate în mod automat în funcție de sarcină și de context. De exemplu, într-un scenariu definit de curând, utilizatorul trebuie să atingă o destinație (ușa IRIT) fiind ghidat pe un traseu de aproximativ 500 m. Poziționarea utilizatorului este ameliorată de detectarea unor ținte geolocalizate de-a lungul parcursului. Modul de localizare a obiectelor este activat în mod automat atunci când utilizatorul se apropie de trecerea de pietoni. Trecerea de pietoni este localizată prin vederea integrată, ceea ce permite o ghidare mai fină. În prezent, se lucrează la conceperea protocoalelor de experimentare care vor permite evaluarea ansamblului funcțiilor implementate în prototip.

Un alt ansamblu de rezultate importante are în vedere percepția prin simțul tactil a unor desene și hărți în relief. Datorită unor experiențe în psihologia experimentală, s-a putut pune în evidență o absență a deficitelor perceptive la persoanele cu deficiențe vizuale și capacitatea acestora de a recunoaște imagini cu semnificație (Samuel Lebaz¹³⁵ et al. 2010; Delphine Picard¹³⁶ et al. 2010; Delphine Picard¹³⁷ et al. 2011; Samuel Lebaz¹³⁸ et al. 2011). Lucrările actuale se referă la concepția participativă a unui prototip interactiv care permite explorarea hărților și pregătirea deplasărilor (Anke Brock¹³⁹ et al. 2010a). Acest prototip va servi ca platformă pentru evaluarea unor noi metode de interacțiune multimodale nevizuale. Metodele de ghidare și dobândirea competențelor spațiale vor fi evaluate comparând diferite tehnici de interacțiune cu dispozitivul. Această abordare a supleenței spațiale prin localizare s-a aplicat în cadrul interfețelor creier-mașină. Experiențele realizate cu simulatorul au arătat că utilizatorii sunt capabili să realizeze sarcini de localizare și de apucare a

¹³³ Kammoun, Slim, Dramas Florian, Bernard Oriola & Christophe Jouffrais, *Route selection algorithm for Blind pedestrian*, Paper presented to the International Conference on Control, Automation and Systems, Gyeonggi-do, South Korea, 2010.

¹³⁴ Brilhault, Adrien, Slim Kammoun, Olivier Gutierrez, Philippe Truillet & Jouffrais Christophe, *Fusion of Artificial Vision and GPS to Improve Blind Pedestrian Positioning*, Paper presented at the 4th IFIP International Conference on New Technologies, Mobility and Security, Paris, France, 2011.

¹³⁵ Lebaz, Samuel, Delphine Picard & Christophe Jouffrais, *Haptic Recognition of Non-figurative Tactile Pictures in the Blind: Does Life-Time Proportion without Visual Experience Matter?*, Amsterdam, NL, 2010.

¹³⁶ Picard, Delphine, Samuel Lebaz, Christophe Jouffrais & Catherine Monnier, *Haptic recognition of two-dimensional raised-line patterns by early-blind, late-blind, and blindfolded sighted adults*, *Perception* 39.224-35, 2010.

¹³⁷ Picard, Delphine, Christophe Jouffrais & Samuel Lebaz, *Haptic Recognition of Emotions in Raised-Line Drawings by Congenitally Blind and Sighted Adults*, *Haptics*, IEEE Transactions on 4.67-71, 2011.

¹³⁸ Lebaz, Samuel, Christophe Jouffrais & Delphine Picard, *Haptic identification of raised-line drawings: high visuospatial imagers outperform low visuospatial imagers*, *Psychol Res.* 1-9, 2011.

¹³⁹ Brock, Anke, Philippe Truillet, Bernard Oriola & Christophe Jouffrais, *Usage of multimodal maps for blind people: why and how*, Paper presented to the ITS'10, 2010a.

obiectelor mult mai repede decât cu o abordare tradițională de tipul „scoreboard”. Aceste rezultate capătă importanță știind că mai multe programe de implantare de retine artificiale sunt în curs în prezent, inclusive în Franța. Abordarea permite utilizarea implanturilor într-un mod mult mai util și utilizabil în sarcini de localizare și navigare decât prin abordarea tradițională.



Fig. 21. Prototip actual asigură două funcții de ajutor: localizarea de obiecte apropiate și poziționarea. (figură preluată din Brian F.G. Katz et al, VRJ 2012).

Figura 21 prezintă un dispozitiv care integrează camere stereoscopice cu o centrală inerțială, un GPS cu o centrală inerțială și un ansamblu de interacțiuni intrare/ieșire bazate pe voce și sunet.

6. Proiecte de cercetare în dezvoltare

Cercetarea întreprinsă în acești ultimi ani a permis validarea ipotezei de localizare a obiectelor ca bază pentru supleanța spațială. S-a demonstrat că cele două tipuri de supleanță avute în vedere (prin realitate mărită și prin interacțiune creier-mașină) sunt funcționale. S-au prezentat, în special, primele rezultate și prototipurile unui dispozitiv de asistare pentru navigare care include un sistem de vedere artificială (camere + algoritm de recunoaștere și localizare a obiectelor), o arhitectură distribuită care permite fuzionarea oricăror tipuri de date în intrare și un sistem de sinteză binaurală în ieșire care permite mărirea obiectelor în spațiu cu informații sonore virtuale. Această abordare, materializată printr-un dispozitiv experimental, este astăzi baza unui proiect de colaborare (NAVIG), finanțat de ANR și etichetat de polul de competitivitate Aerospace Valley¹⁴⁰. În secțiunea următoare, vor fi descrise piedicile apărute în derularea acestei cercetări, la care se lucrează în prezent și în următorii doi ani.

Pe parcursul avansării acestei cercetări cu privire la asistarea navigării, s-a identificat clar importanța de a concepe un simulator de navigare în realitatea virtuală. Acest simulator ar permite evaluarea tehnicilor de poziționare și conceperea unor metode de interacțiune și de ghidare novatoare, mai rapid, în situații controlate

¹⁴⁰ A se vedea <http://navig.irit.fr>

și fără risc pentru utilizatori¹⁴¹. O altă ipoteză este că aceste metode vor fi transferabile într-o situație reală, in situ.

Al patrulea aspect al cercetării are în vedere explorarea unor hărți interactive multimodale. Este deosebit de important pentru persoanele cu deficiențe vizuale să poată pregăti o deplasare căutând repere și itinerarii adaptate capacităților și tehnicilor lor de mobilitate și de orientare. Se dorește extinderea acestei cercetări către IHM, prin interacțiunea multimodală cu hărți în relief interactive.

De asemenea, este foarte important tematica interfețelor creier-mașină și a simulatorului de neuroproteză conceput. Acest simulator este bazat pe ipoteza principală conform căreia localizarea unor obiecte pertinente în funcție de sarcină și de context permite recuperarea a numeroase comportamente vizual-motorii. Împreună cu Marc Macé, s-a planificat realizarea a numeroase lucrări referitoare la aspectele tehnologice, fundamentale și comportamentale ale neuroprotezilor vizuale.

În colaborare cu Institutul Tinerilor Orbi din Toulouse, se încearcă reunirea cercetătorilor, a persoanelor cu deficiențe vizuale și a formatorilor în jurul unor proiecte comune privind tehnologiile de asistare pentru persoanele cu deficiențe vizuale.

7 Asistare electronică pentru navigare

7.1.1 Descrierea proiectului

Obiectivul final al proiectului este de a concepe un demonstrator interactiv care să permită evaluarea pe loc (la universitate, dar și în centrul orașului) a dispozitivului NAVIG în scenarii clar definite ca problematice pentru persoanele cu deficiențe vizuale. Această evaluare este posibilă datorită participării Institutului Tinerilor Orbi, pe de o parte, și datorită proiectului Toulouse OPEN care pune la dispoziția cercetătorilor toate instrumentele necesare evaluării precise, in situ, a dispozitivului. Pentru obținerea și modificarea bazelor de date geografice, s-a realizat o colaborare cu serviciul SIG al comunității Grand Toulouse. Dintr-un punct de vedere fundamental, acest proiect va permite realizarea unor progrese importante în neuroștiințe (percepție auditivă, acțiuni auditive ghidate, cogniție spațială la nevăzători), dar și în domeniul supleanței pentru nevăzători (interacțiunea om-mașină, întrebuintări și utilizabilitate a sistemelor de supleanță).

7.1.2 Principalele obstacole

Principalele obstacole tehnologice se situează la nivelul vederii artificiale și a poziționării. Obstacolele de ordin științific se situează, pentru neuroștiințe, la nivelul percepției și al cogniției spațiale 3D și al acțiunii ghidate auditiv, iar pentru IHM, la nivelul supleanței funcționale și al interacțiunii degradate.

¹⁴¹ Într-adevăr, este extrem de greu, chiar periculos să se evalueze metode de ghidare în situații reale. Deși s-a dezvoltat metodologia pentru realizarea acestor evaluări, se preferă efectuarea unui ansamblu de teste prealabile în laborator.

Unul dintre obstacolele tehnologice ale modulului de vedere artificială constă în specificarea și conceperea numărului, mărimii și ierarhiei modelelor necesare pentru a ajunge la o fiabilitate și o robustețe de detectare a țintelor (în special 3D) compatibilă cu întrebuințările dispozitivului. Un obstacol de ordin științific constă în studiul și dezvoltarea unui algoritm care permite modificarea, în mod dinamic, a rezoluțiilor de cercetare a modelelor, precum și utilizarea fluxurilor optice pentru detectarea obstacolelor. Aceste aspecte teoretice și aplicarea lor la supleanța pentru nevăzători vor fi adresate în cadrul colaborării între IRIT, CerCo și SpikeNet Technology și, în special, în cadrul tezei lui Adrien Brilhault.

Obstacolele potențiale pentru ghidarea auditivă sunt conflictele între redarea audio și lumea reală, precum și designul ergonomic al paletelor de sunete pentru o utilizare cotidiană. Aceste două aspecte fac să apară constrângeri tehnologice și ergonomice. Pentru componenta tehnologică, LIMSI, partener al proiectului, lucrează la problema echipamentului cel mai bine adaptat pentru realitatea auditivă mărită și se va referi la problemele legate de designul sonor și de ergonomie. În colaborare cu aceștia (echipa lui Brian F.G. Katz și M. Denis), se va încerca o mai bună înțelegere a capacităților perceptive și cognitive ale utilizatorilor, precum și acțiunile care survin de aici. Această cercetare a fost inițiată în cadrul tezei lui Florian Dramas la IRIT și este reluată în cadrul tezei lui Gaétan Parseihan, condusă de Brian F.G. Katz (LIMSI).

La IRIT, în colaborare cu Philippe Truillet, MCU în informatică din echipa ELIPSE, se derulează o cercetare cu privire la fuziunea unor date heterogene (în special poziționarea GPS, centrala inerțială și vederea), luând în considerare constrângerile legate de sistemele informaționale geografice.

Această fuziune poate fi asimilată unei fuziuni „multimodale” în care datele corespund unor fluxuri concurente și/sau complementare care diferă prin natura, frecvența de eșantionare și semantica lor. Se va încerca să se demonstreze că fuziunea acestor date permite ameliorarea în mod semnificativ a poziționării pietonilor nevăzători în medii zgomotoase (de exemplu, orașul) și, prin urmare, creșterea autonomiei lor în sarcini de navigare (de exemplu, a merge la poșta din centrul orașului și a găsi cutia poștală).

Această cercetare este în curs în cadrul tezei lui Slim Kammoun, condusă de Christophe Jouffrais, cu ajutorul unui inginer contractual (O. Gutierrez), finanțată prin proiectul NAVIG.

7.2 Simulator de navigare

EOA (ajutoare electronice pentru orientare) sunt dedicate asistării orientării și sunt constituite din trei elemente esențiale:

- 1) un sistem de poziționare (GPS de exemplu);
- 2) un sistem informațional geografic (SIG) care cuprinde, în același timp, o hartă digitală și softuri concepute pentru a selecta itinerare, a urmări traseul călătorului și a-i furniza informații pentru navigare;
- 3) o interfață utilizator care are la bază o interacțiune nevizuală (în general auditivă sau tactilă).

Toate aceste trei componente sunt la originea unor probleme legate de eficiența utilizării. Primul obstacol are în vedere eroarea de poziționare GPS care este deseori mai mare de 10 metri, ceea ce nu este compatibil cu ajutorul pentru navigare pentru persoanele cu deficiențe vizuale.

Al doilea obstacol are în vedere SIG care, în general, conțin exclusiv informații care privesc rețelele rutiere și nici o informație relativă la zonele pietonale. În sfârșit, ultimul obstacol are în vedere interacțiunea între utilizator și sistem și interacțiunea de ghidare. Acestea trebuie concepute în cadrul unei colaborări cu persoanele cu deficiențe vizuale și cu formatorii în domeniul locomoției.

Pentru a accelera procesele de concepție, mediile virtuale interactive reprezintă o platformă prețioasă pentru a izola în mod selectiv diferitele componente și a le evalua în mod sistematic în laborator înainte de a fi evaluate la fața locului.

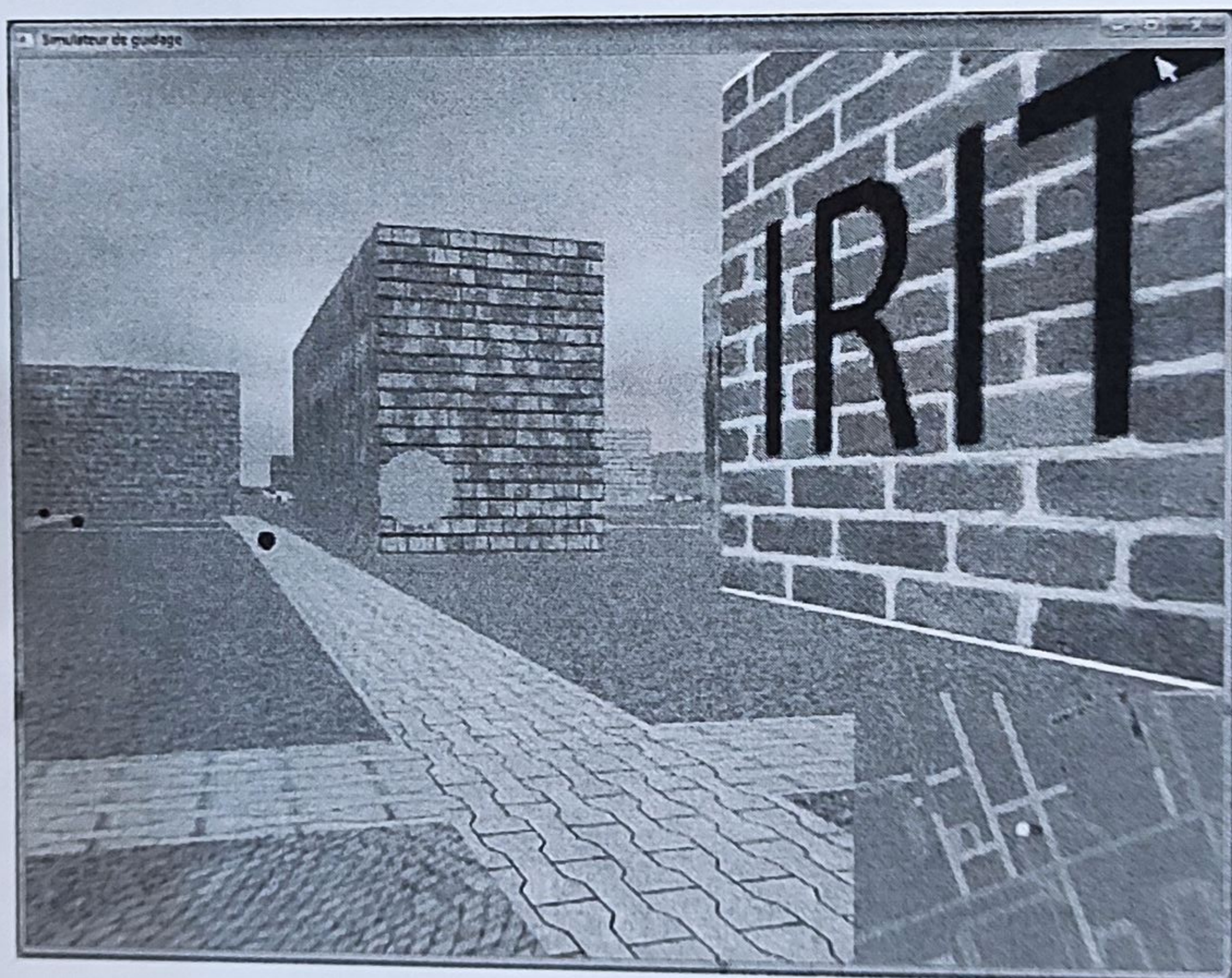


Fig. 22. Soft dedicat de simulre a mediului virtual - programat pentru zona de amplasare a institutului IRIT (instantiere din programul conceput la IRIT)

S-a demarat dezvoltarea unei platforme experimentale care să permită persoanelor cu deficiențe vizuale să exploreze un mediu virtual prin interacțiuni auditive și haptice. Această platformă permite utilizatorului să navigheze într-un mediu virtual care reprezintă un spațiu real (bazat pe SIG al unei universități sau al unui oraș, de exemplu). Obiectivul acestei platforme este de a testa în mod sistematic procesele de ghidare înainte de punerea lor în aplicare la fața locului, în cadrul unui EOA. Aceste teste vor permite să se determine ce procese de ghidare sunt cele mai eficiente pentru a compensa o poziționare GPS greșită, în același timp ameliorând dobândirea de cunoștințe spațiale.

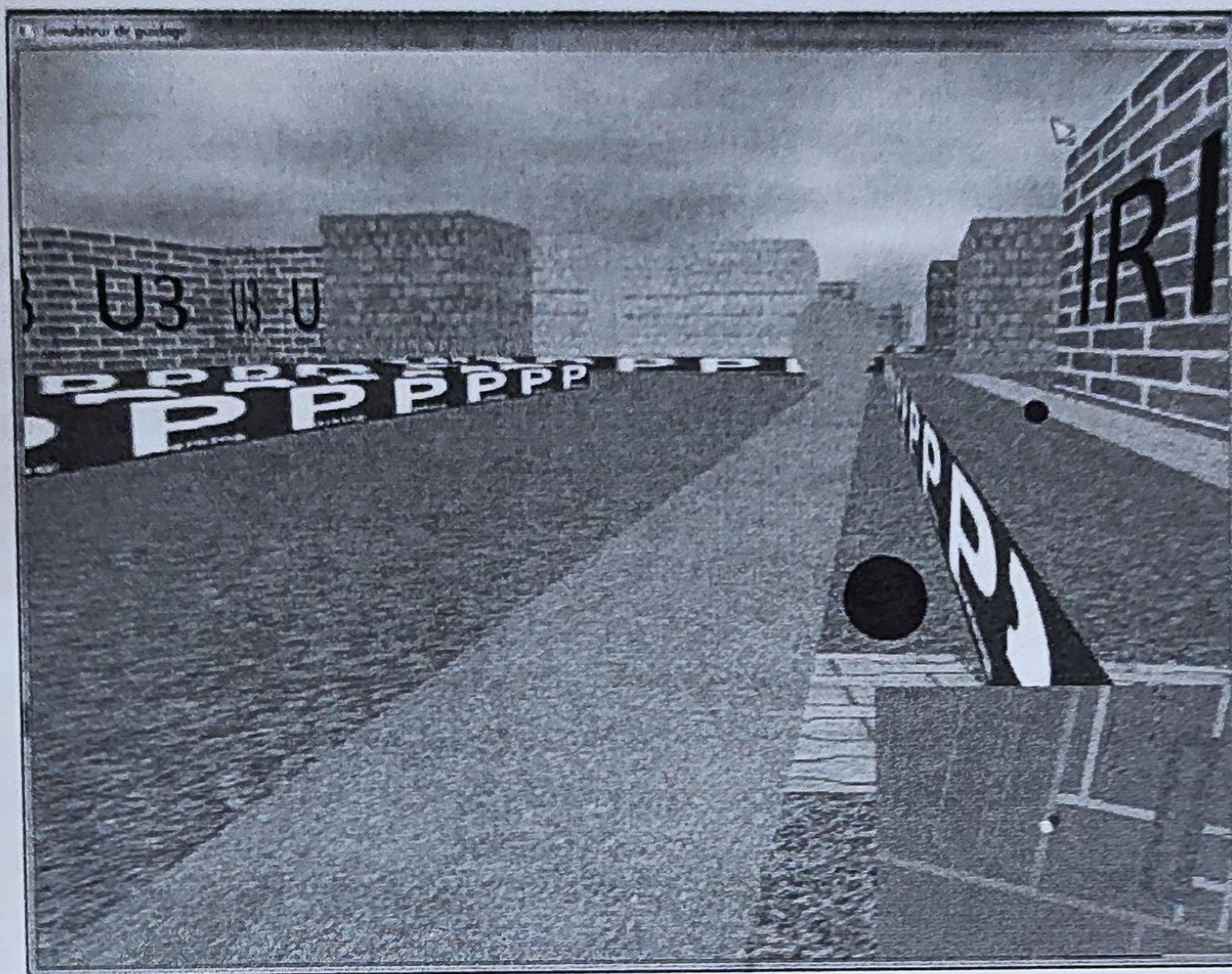


Fig. 23 *Deplasare în zona IRIT, UPS, Toulouse - mediul virtual simulat prin soft*

7.3 Hărți interactive multimodale

În continuarea cercetărilor cu privire la percepția hărților prin simțul tactil, s-au prevăzut două noi axe de lucru: conceperea de noi interacțiuni în intrare care permit explorarea hărților tactile și conceperea unui sistem de ecran deformabil care permite afișarea hărților dinamice și oferirea unui feed-back tactil.

7.3.1 Conceperea unor noi metode de interacțiune spațială

Lucrarea de analiză realizată în colaborare cu IJA în cadrul tezei lui Anke Brock a permis definirea mai multor funcții care ar fi utile pentru a permite persoanelor cu deficiențe vizuale să dobândească cunoștințe spațiale indispensabile înainte de a efectua o deplasare. Un prim ansamblu de funcții are în vedere asistarea planificării unui traseu, evitând zonele desagrabile sau periculoase și incluzând zonele preferate. Un al doilea ansamblu de funcții are în vedere asistarea învățării configurației unor zone speciale, incluzând obiectele importante care intră în constituția sa.

Va fi necesară conceperea, împreună cu utilizatorii, a unui ansamblu de noi interacțiuni care să permită implementarea tuturor acestor funcții, apoi evaluarea lor din perspectiva eficienței utilizării, dar și a asistenței procesului de învățare spațială. Se prevede și evaluarea transferului acestor cunoștințe atunci când învățarea pe hartă este urmată de o deplasare la fața locului.

7.3.2 Redarea informațiilor tactile

Tehnologiile tactile actuale se bazează, în principal, pe interacțiuni grafice sau sonore (a se vedea tabletele tactile, cu o evoluție crescândă). Această concepție duce la situații și mai ironice pentru persoanele cu deficiențe vizuale, care nu mai au niciun reper tactil pentru a interacționa cu dispozitivul.

Ne gândim, în colaborare cu școala de ingineri ESTIA din Bayonne, la un ecran tactil deformabil care ar permite conceperea unor interacțiuni accesibile pentru persoanele cu deficiențe vizuale. Obiectivul este, în special, de a putea afișa hărți dinamice, ceea ce ar permite, de exemplu, zoom-ul și defilarea, precum și feed-back-uri în cursul diferitelor interacțiuni. Se urmează astfel o a doua pistă, în colaborare cu Universitatea din Lille (a se vedea Casiez¹⁴² et al. 2011), care ar permite utilizarea unui feed-back vibrator. Această idee s-ar putea sprijini pe existența a numeroase interfețe tactile sau haptice existente și adaptate pentru a fi utilizate de către persoanele cu deficiențe vizuale (a se vedea, de exemplu, Velázquez & Pissaloux¹⁴³ 2008).

7.4 Interfețele neurale și neuroprotezele vizuale

7.4.1 Evocarea unor fosfene localizate și având semnificație, ca intrare pentru comportamentele vizual-motorii

În colaborare cu CerCo (neuroștiințe) Centre de Recherche Cerveau & Cognition – Université Toulouse III, LAAS (nanotehnologii) Laboratoire d'Analyse et d'Architecture des Systèmes, și clinica de medicină fizică și reeducare din cadrul CHU Rangueil, Centre Hospitalier Universitaire, lucrăm la aspectele tehnologice, fundamentale și comportamentale ale neuroprotezelor vizuale.

Din punct de vedere tehnologic, derulăm o cercetare referitoare la interfețele neurale, la biocompatibilitatea acestora și la raportul semnal/zgomot. Această cercetare tehnologică vine în completarea lucrărilor efectuate împreună cu Denis Fize la CerCo (Denis Fize¹⁴⁴ et al. 2009), care au drept obiectiv dezvoltarea unui model de neuroproteză vizuală pentru oameni. Pe termen mediu, se prevede testarea protocoalelor de interacțiune definite la IRIT pe animale implantate cu interfețe dezvoltate la LAAS. În final, la IRIT, se acordă o mare importanță cercetării privind simularea unei neuroproteze vizuale de localizare a obiectelor, încercându-se să se demonstreze că această abordare permite recuperarea a numeroase comportamente vizual-motorii fundamentale.

¹⁴² Casiez, Géry, Nicolas Roussel, Romuald Vanbelleghem & Frédéric Giraud, *Surfpad: riding towards targets on a squeeze film effect*, Paper presented at the CHI 2011, Vancouver, BC, Canada, 2011.

¹⁴³ Velázquez, Ramiro & Edwige E. Pissaloux, *Tactile displays in human-machine interaction: four case studies*, International Journal of Virtual Reality 7.51-58, 2008.

¹⁴⁴ Fize, Denis, Maxime Cauchoux & Christophe Jouffrais, *Visual scene processing Dynamics in Monkeys: building the information content of a visual neuroprosthesis*, Paper presented at the Congrès FRM/IRME/ICM 2009 sur les Interfaces Cerveau-Machine, Paris, France, 2009.

Pentru a dezvolta o alternativă la abordarea de tip „scoreboard” și a face dovada conceperii unei neuroproteze de localizare a obiectelor, a fost dezvoltat, la IRT, un prototip și un post experimental care permite conceperea unui ansamblu de experiențe comportamentale ce fac să varieze numărul, poziția și spațiarea fosfenelor. Intensitatea luminoasă va fi, de asemenea, un parametru disponibil (Girvin¹⁴⁵ et al. 1979) pentru a da o informație de profunzime, de exemplu. Această abordare, bazată pe localizarea unor ținte, va fi comparată în mod sistematic cu abordarea mai clasică de tipul „scoreboard”. Precizia comportamentelor în spațiu va fi măsurată fie în 2D prin imagini video, fie în 3D cu ajutorul unui dispozitiv de captură a mișcării. Acest studiu va fi realizat în cadrul tezei lui Grégoire Denis, sub responsabilitatea lui Marc Macé, în colaborare cu echipa SC a IRT și cu INSERM U825.

Această experimentare ar trebui să arate că o neuroproteză vizuală, bazată pe o rezoluție compatibilă cu implanturile actuale (cel mult câteva sute de electrozi), având ca principiu recunoașterea și localizarea obiectelor de interes, permite construirea unei reprezentări „rarefiate” dar funcționale a spațiului, care poate fi utilizată pentru a produce comportamente orientate către obiecte sau locuri. Această abordare este transversală abordării clasice de tipul „scoreboard”. Dacă validăm ipotezele emise, vom căuta să punem în practică, pe pacienți implantați, modalitățile noastre de interacțiune bazate pe localizarea de obiecte. Acest lucru s-ar putea face în cadrul unei colaborări, prin una din încercările terapeutice puse în aplicare în prezent, referitoare la implantarea de retine artificiale la om¹⁴⁶.

Bibliografie

- Adams, Christi J., *An investigation of navigation processes in human locomotion behavior*, Master of science Virginia: Polytechnic Institute and State University, 1997.
- Afonso, Amandine, Brian F.G. Katz, Alan Blum & Michel Denis, *Mental imagery and the acquisition of spatial knowledge without vision: A study of blind and sighted people in an immersive audio virtual environment*, St Andrews, Scotland, 2005.
- Auvray, Malika, Sylvain Hanneton & J. Kevin O'Regan, *Learning to perceive with a visuo-auditory substitution system: Localisation and object recognition with The vOICE*, Perception 36.416, 2007.
- Arno, Patricia, Annick Vanlierde, Emmanuel Streel, Marie-Chantal Wanet-Defalque, Sandra Sanabria-Bohorquez & Claude Veraart, *Auditory substitution of vision: pattern recognition by the blind*, Applied Cognitive Psychology 15.509-19, 2001.
- Bach-y-Rita, Paul, Carter C. Collins, Benjamin White, Frank A. Saunders, Lawrence Scadden & Robert Blomberg, *A tactile vision substitution system*, Am.J.Optom.Arch.Am.Acad.Optom. 46.109-11, 1969.
- Bailes, Sally M. & Robert M. Lambert, *Cognitive aspects of haptic form recognition by blind and sighted subjects*, British Journal of Psychology 77.451-58, 1986.
- Barber, Paul O. & Susan J. Lederman, *Encoding direction in manipulatory space and the role of visual experience*, Journal of Visual Impairment and Blindness 82.99-106, 1988.
- Bellik, Yacine & René Farcy, *Comparison of Various Interface Modalities for a Locomotion Assistance Device*, Computers Helping People with Special Needs, ed. by K. Miesenberger, J. Klaus & W. Zagler, 125-36: Springer Berlin / Heidelberg, 2002.
- Bentzen, Billie Louise & Philip A. Mitchell, *Audible Signage as a Wayfinding Aid: Verbal Landmark versus Talking Signs*, Journal of Visual Impairment & Blindness 89.494-505, 1995.
- Blauert, Jens, *Spatial hearing: the psychophysics of human sound localization*, MIT Press, 1997.

¹⁴⁵ Girvin, John P., Jerald R. Evans, William H. Doherty, Michael G. Mladejovsky, David C. Henderson, Israel Abramov, James Gordon & Joseph Turkel, *Electrical stimulation of human visual cortex: the effect of stimulus parameters on phosphene threshold*, Sens.Processes 3.66-81, 1979.

¹⁴⁶ Lucrări în curs la Institut de la Vision, Paris: http://www.institut-vision.org/index.php?view=article&id=101:des-implants-pour-vaincre-la-cessite&option=com_content&Itemid=16&lang=fr

CHRISTOPHE JOUFFRAIS

- Borovec, Jiri, *Fusion of heterogeneous data for better positioning of visually impaired pedestrians*, Master degree: Université de Toulouse, 2011.
- Brabyn, Lesley A. & John A. Brabyn, *An evaluation of talking signs for the blind*. *Human Factors*, The Journal of the Human Factors and Ergonomics Society 25.49-53, 1983.
- Brambring, Michael, *The structure of haptic space in the blind and sighted*, *Psychological Res.* 38.283-302, 1976.
- Brilhault, Adrien, Slim Kammoun, Olivier Gutierrez, Philippe Truillet & Jouffrais Christophe, *Fusion of Artificial Vision and GPS to Improve Blind Pedestrian Positioning*, Paper presented at the 4th IFIP International Conference on New Technologies, Mobility and Security, Paris, France, 2011.
- Brindley, Giles S., *Effects of electrical stimulation of the visual cortex*, *Hum. Neurobiol.* 1.281-83, 1982.
- Brindley, Giles S., Peter E. K. Donaldson, Murray A. Falconer & David N. Rushton, *The extent of the region of occipital cortex that when stimulated gives phosphenes fixed in the visual field*, *J. Physiol* 225.57P-58P, 1972.
- Brindley, Giles S. & Walpole S. Lewin, *The sensations produced by electrical stimulation of the visual cortex*, *J. Physiol* 196.479-93, 1968.
- Brock, Anke, *Cartes interactives multimodales pour piétons malvoyants*, Master thesis UPS, Toulouse, 2010.
- Brock, Anke, Philippe Truillet, Bernard Oriola & Christophe Jouffrais, *Usage of multimodal maps for blind people: why and how*, Paper presented to the ITS'10, 2010a.
- Brock, Anke, Jean-Luc Vinot, Bernard Oriola, Slim Kammoun, Philippe Truillet & Christophe Jouffrais, *Méthodes et outils de conception participative avec des utilisateurs non-voyants*, Paper presented to the IHM'10, 2010b.
- Bruce, Ian, Aubrey McKennel & Errol Walker, *Blind and partially sighted adults in Britain: The RNIB Needs Survey*, Volume 1. HMSO, London, 1991.
- Buisson, Marcellin, Alexandre Bustico, Stéphane Chatty, François-Régis Colin, Yannick Jestin, Sébastien Maury, Christophe Mertz & Philippe Truillet, *Ivy: un bus logiciel au service du développement de prototypes de systèmes interactifs*, 2002.
- Burrough, Peter A., *Principles of geographical information systems for land resources assessment*, Clarendon Press, 1994.
- Caddeo, Pierluigi, Ferdinando Fornara, Anna Maria Nenci & Amelia Piroddi, *Wayfinding tasks in visually impaired people: the role of tactile maps*, *Cognitive Processing* 7.168-69, 2006.
- Capelle, Christian, Charles Trullemans, Patricia Arno & Claude Veraart, *A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution*, *Ieee Transactions on Biomedical Engineering* 45.1279-93, 1998.
- Casey, Steven M., *Cognitive mapping by the blind*, *Journal of Visual Impairment & Blindness* 72.297-301, 1978.
- Casiez, Géry, Nicolas Roussel, Romuald Vanbelleghem & Frédéric Giraud, *Surfpad: riding towards targets on a squeeze film effect*, Paper presented at the CHI 2011, Vancouver, BC, Canada, 2011.
- Clark-Carter, David D., Anthony D. Heyes & Ian Howarth, *The efficiency and walking speed of visually impaired people*, *Ergonomics* 29, 1986.
- Dejeammes, Maryvonne, Gérard Uzan, M'Balo Seck & Catherine Sidot, *Déplacements des déficients visuels en milieu urbain: analyse des besoins en sécurité, localisation et orientation et pistes d'évolution*, CERTU, 2008.
- Dijkstra, Edsger W., *A note on two problems in connexion with graphs*, *Numerische mathematik* 1.269-71, 1959.
- Dobelle, William H. & Michael G. Mladejovsky, *Phosphenes produced by electrical stimulation of human occipital cortex, and their application to the development of a prosthesis for the blind*, *J. Physiol* 243.553-76, 1974.
- Dobelle, William H., Michael G. Mladejovsky & John P. Girvin, *Artificial vision for the blind: electrical stimulation of visual cortex offers hope for a functional prosthesis*, *Science* 183.440-44, 1974.
- Dodds, Allan G., Ian Howarth & David Clark-Carter, *The mental maps of the blind: The role of previous visual experience*, *Journal of Visual Impairment & Blindness* 76.5-12, 1982.
- Dowling, Jason, *Mobility enhancement using simulated artificial human vision*, PhD: Queensland University of Technology, 2003.
- Downs, Roger M. & David Stea, *Image & environment: Cognitive mapping and spatial behavior*, Edward Arnold, London, 1973.
- Downs, Roger M. & David Stea, *Maps in minds: reflections on cognitive mapping*, Harper & Row, 1977.
- Dramas, Florian, *Localisation d'objets pour les non-voyants : augmentation sensorielle et neuroprothèse*, PhD: University of Toulouse 3, France, 2010.
- Dramas, Florian, Brian F.G. Katz & Christophe Jouffrais, *Auditory-guided reaching movements in the peripersonal frontal space*, Paper presented at the The Journal of the Acoustical Society of America, 2008.
- Dramas, Florian, Simon J. Thorpe & Christophe Jouffrais, *Localisation d'objets pour les non-voyants par analyse d'image: analyse du besoin et prototypage*, Paper presented to the ASSISTH, Toulouse, FR, 2007.
- Dramas, Florian, Simon J. Thorpe & Christophe Jouffrais, *Artificial Vision For The Blind: A Bio-Inspired Algorithm For Objects And Obstacles Detection*, *International Journal of Image and Graphics* 10.531-44, 2010.
- Durette, Barthélémy, *Traitement du signal pour les prothèses visuelles: approche biométrique et sensori-motrice*, PhD: Université Grenoble I, 2009.
- Durette, Barthélémy, Nicolas Louveton, David Alleysson & Jeanny Hérault, *Visuo-auditory sensory substitution for mobility assistance: testing TheVIBE*, Paper presented at the CVAVI'08, Marseille, France, 2008.
- Eriksson, Yvonne, *Tactile pictures: pictorial representations for the blind*, *Acta Universitatis Gothoburgensis*, 1998.

Tehnologii de asistare a persoanelor cu deficiențe vizuale

- Farcy, René & Roland M. Damaschini, *Guidance - Assist systems for the blind*, Paper presented at the Proceedings of the Biomonitoring and Endoscopy Technologies, 2001.
- Farcy, René, Roger Leroux, Roland M. Damaschini, Richard Legras, Yacine Bellik, Christophe Jacquet, Janet Greene & Pierre Pardo, *Laser telemetry to improve the mobility of blind people: Report of the 6 month training course*, Paper presented at the ICOST, 2003.
- Farcy, René, Roger Leroux, Alain Jucha, Roland M. Damaschini, Colette Gregoire & Aziz Zogaghi, *Electronic travel aids and electronic orientation aids for blind people: technical, rehabilitation and everyday life points of view*, Paper presented at the CVHI, 2006.
- Fish, Raymond M., *An audio display for the blind*, Biomedical Engineering, IEEE Transactions on.144-54, 1976.
- Fize, Denis, Maxime Cauchoux & Christophe Jouffrais, *Visual scene processing Dynamics in Monkeys: building the information content of a visual neuroprosthesis*, Paper presented at the Congrès FRM/IRME/ICM 2009 sur les Interfaces Cerveau-Machine, Paris, France, 2009.
- Fletcher, Janet F., *Spatial Representation in Blind Children. 1: Development Compared to Sighted Children*, Journal of Visual Impairment and Blindness 74.381-85, 1980.
- Foulke, Emerson, *The perceptual basis for mobility*, American Foundation for the Blind, Research Bulletin, 1971.
- Foulke, Emerson, *Perception, cognition, and the mobility of blind pedestrians*. Spatial abilities: Development and physiological foundations, ed. by M. Potegal, 55-76: New York: Academic Press, 1982.
- Fujiki, Nobuya, Klaus A. Riederer, Veikko Jousmaki, Jyrki P. Makela & Riitta Hari, *Human cortical representation of virtual auditory space: differences between sound azimuth and elevation*, Eur.J.Neurosci. 16.2207-13, 2002.
- Gallagher, Blathin & Walter Frasc, *Tactile Acoustic Computer Interaction System: A new type of Graphic Access for the Blind*, Paper presented at the 3rd TIDE congress 1998, Helsinki, 1998.
- Gallistel, Charles R., *The organization of learning*, The MIT Press, 1990.
- Garling, Tommy & Reginald G. Golledge, *Cognitive mapping and spatial decision-making*. Cognitive mapping past, present and future, ed. by R.M. Kitchin & S. Freundschuh, 44-65: Routledge, 2001.
- Girvin, John P., Jerald R. Evans, William H. Dobelle, Michael G. Mladejovsky, David C. Henderson, Israel Abramov, James Gordon & Joseph Turkel, *Electrical stimulation of human visual cortex: the effect of stimulus parameters on phosphene threshold*, Sens.Processes 3.66-81, 1979.
- Golledge, Reginald G., *Tactual strip maps as navigational aids*, J. Visual Impairment&Blindness 85.296-301, 1991.
- Golledge, Reginald G., *Geography and the disabled: a survey with special reference to vision impaired and blind populations*, Transactions of the Institute of British Geographers 18.63-85, 1993.
- Golledge, Reginald G., *Wayfinding behavior: cognitive mapping and other spatial processes*, JHU Press, 1999.
- Golledge, Reginald G., Matthew Rice & Daniel R. Jacobson, *A Commentary on the Use of Touch for Accessing On-Screen Spatial Representations: The Process of Experiencing Haptic Maps and Graphics*, The Professional Geographer 57.339-49, 2005.
- Golledge, Reginald G. & Robert J. Stimson, *Analytical behavioural geography*, Guilford Press, 1997.
- Guivarch, Valérian, *Modélisation d'une neuroprothèse de reconnaissance et de localisation d'objet dans l'espace*, Master IT-ICSC: University of Toulouse 3, 2010.
- Hatsopoulos, Nicholas G. & John P. Donoghue, *The science of neural interface systems*, Annual Review of Neuroscience 32.249-66, 2009.
- Helal, Abdelsalam, Steven Edwin Moore & Balaji Ramachandran, *Drishti: An integrated navigation system for visually impaired and disabled*, Paper presented at the ISWC'01, Zurich, Switzerland, 2001.
- Heller, Morton A., *Picture and pattern perception in the sighted and the blind: the advantage of the late blind*, Perception 18.379-89, 1989.
- Heller, Morton A., *Tactile picture perception in sighted and blind people*, Behavioural Brain Res. 135.65-68, 2002.
- Heller, Morton A. & John M. Kennedy, *Perspective taking, pictures, and the blind*, Percept Psychophys 48.459-66, 1990.
- Heller, Morton A., Melissa McCarthy & Ashley Clark, *Pattern perception and pictures for the blind*, Psicológica 26.161-71, 2005.
- Heuten, Wilko, Daniel Wichmann & Susanne Boll, *Interactive 3D sonification for the exploration of city maps*, Paper presented to the NordiCHI 2006, 2006.
- Hinton, Ron, *Tactile and audio-tactile images as vehicles for learning*, Paper presented to the Non-visual human-computer interactions: prospects for the visually handicapped, Paris, 1993.
- Hollyfield, Rebecca L. & Emerson Foulke, *The spatial cognition of blind pedestrians*, Journal of Visual Impairment & Blindness, 1983.
- Jacobson, Daniel R., *Navigating maps with little or no sight: An audio-tactile approach*, Paper presented at the Content Visualization and Intermedia Representations, University of Montreal, Montreal, 1998.
- Juurmaa, Jyrki, *Transportation in Mental Spatial Manipulation: A Theoretical Analysis*, American Foundation for the Blind Research Bulletin 26.87-143, 1973.
- Kammoun, Slim, Dramas Florian, Bernard Oriola & Christophe Jouffrais, *Route selection algorithm for Blind pedestrian*, Paper presented to the International Conference on Control, Automation and Systems, Gyeonggi-do, South Korea, 2010.
- Katz, Brian F. G. & Lorenzo Picinali, *Spatial Audio Applied to Research with the Blind*, Advances in Sound Localization, ed. by P. Strumillo: InTech, 2011.

CHRISTOPHE JOUFFRAIS

- Kennedy, John M., *Drawing and the blind: pictures to touch*, Yale University Press, 1993.
- Klatzky, Roberta L., Susan J. Lederman & Victoria A. Metzger, *Identifying objects by touch: an "expert system"*, *Percept Psychophys* 37, 1985.
- Klatzky, Roberta L., Yvonne Lippa, Jack M. Loomis & Reginald G. Golledge, *Learning directions of objects specified by vision, spatial audition, or auditory spatial language*, *Learning and Memory* 9.364-67, 2002.
- Kulkarni, Abhijit & H. Steven Colburn, *Role of spectral detail in sound-source localization*, *Nature* 396.747-49, 1998.
- Lebaz, Samuel, Christophe Jouffrais & Delphine Picard, *Haptic identification of raised-line drawings: high visuospatial imagers outperform low visuospatial imagers*, *Psychol Res.* 1-9, 2011.
- Lebaz, Samuel, Delphine Picard & Christophe Jouffrais, *Haptic Recognition of Non-figurative Tactile Pictures in the Blind: Does Life-Time Proportion without Visual Experience Matter?*, Amsterdam, NL, 2010.
- Lederman, Susan J. & Roberta L. Klatzky, *Hand movements: a window into haptic object recognition*, *Cogn Psychol* 19.342-68, 1987.
- Lederman, Susan J., Roberta L. Klatzky, Cynthia Chataway & Craig D. Summers, *Visual mediation and the haptic recognition of two-dimensional pictures of common objects*, *Percept Psychophys* 47.54-64, 1990.
- Likert, Rensis & William H Quasha, *Minnesota Paper Form Board Test*, The Psychological Corporation, 1941.
- Loomis, Jack M., Reginald G. Golledge & Roberta L. Klatzky, *Navigation System for the Blind: Auditory Display Modes and Guidance*, *Presence-Teleoperators and Virtual Environments* 7.193-203, 1998.
- Loomis, Jack M., Reginald G. Golledge, Roberta L. Klatzky, Jon M. Speigle & Jerome Tietz, *Personal guidance system for the visually impaired*, Marina Del Rey, California, United States, 1994.
- Loomis, Jack M., Roberta L. Klatzky, Reginald G. Golledge, Joseph G. Cicinelli, James W. Pellegrino & Phyllis A. Fry, *Nonvisual navigation by blind and sighted: assessment of path integration ability*, *J.Exp.Psychol.Gen.* 122.73-91, 1993.
- Meijer, Peter B. L., *An experimental system for auditory image representations*, *IEEE Trans.Biomed.Eng* 39.112-21, 1992.
- Middlebrooks, John C. & David M. Green, *Sound localization by human listeners*, *Annu.Rev.Psychol.* 42.135-59, 1991.
- Millar, Susanna, *Spatial representation by blind and sighted children*, *J.Exp.Child Psychol* 21.460-79, 1976.
- Millar, Susanna, *Self-referent and movement cues in coding spatial location by blind and sighted children*, *Perception* 10.255-64, 1981.
- Millar, Susanna, *Studies of the deaf and the blind*, The Pathology and psychology of cognition, 135: Taylor & Francis, 1982.
- Millar, Susanna, *Understanding and representing space: Theory and evidence from studies with blind and sighted children*, Oxford University Press, USA, 1994.
- Norman, Donald A. & Stephen W. Draper, *User centered system design* Hillsdale, NJ, 1986.
- Parlouar, Rémi, Florian Dramas, Marc M. Macé & Christophe Jouffrais, *Assistive device for the blind based on object recognition: an application to identify currency bills*, Paper presented to the ASSETS'09, Pittsburgh, USA, 2009.
- Passini, Romedi & Guylène Proulx, *Wayfinding without vision*, *Environment and Behavior* 20.227-52, 1988.
- Petrie, Helen, Valerie Johnson, Thomas Strothotte, Andreas Raab, Steffi Fritz & Rainer Michel, *MoBIC: Designing a travel aid for blind and elderly people*, *Journal of Navigation* 49.45-52, 1996.
- Picard, Delphine, Christophe Jouffrais & Samuel Lebaz, *Haptic Recognition of Emotions in Raised-Line Drawings by Congenitally Blind and Sighted Adults*, *Haptics, IEEE Transactions on* 4.67-71, 2011.
- Picard, Delphine, Samuel Lebaz, Christophe Jouffrais & Catherine Monnier, *Haptic recognition of two-dimensional raised-line patterns by early-blind, late-blind, and blindfolded sighted adults*, *Perception* 39.224-35, 2010.
- Ran, Lisa, Sumi Helal & Steve Moore, *Drishti: an integrated indoor/outdoor blind navigation system and service*, Paper presented to the Second IEEE International Conference on Pervasive Computing and Communications (PerCom'04), 2004.
- Révész, Géza, *Psychology and art of the blind* Oxford, England Longmans, Green, 1950.
- Rieser, John J., David A. Guth & Everett W. Hill, *Sensitivity to perspective structure while walking without vision*, *Perception* 15.173-88, 1986.
- Roder, Brigitte & Helen Neville, *Developmental functional plasticity. Handbook of neuropsychology*, ed. by J. Grafman & I.H. Robertson, 231-70: Elsevier, 2003.
- Sander, Marie-Sylvie, Marie-Christine Bournot, Françoise Lelièvre & Anne Tallec, *Les personnes ayant un handicap visuel. Apports de l'enquete handicaps - incapacités - dépendances* (ed.) R.d.e.d.M.F.d.I.S.e.d. Solidarités, 2005.
- Simson, Helen, Deborah Gold & Biljana Zuvella, *Inégalité des chances: Rapport sur les besoins des personnes aveugles ou handicapées visuelles vivant au Canada*, 2005.
- Smyth, Mary M., Alan M. Wing, *The Psychology of human movement*, Academic Press, 1984.
- Strelow, Edward W., *What is needed for a theory of mobility: Direct perceptions and cognitive maps—lessons from the blind*, *Psychological review* 92.226, 1985.
- Tatham, Andrew F. & Allan G. Dodds, *Proceedings of the Second International Symposium on Maps and Graphics for Visually Handicapped People*, Paper presented to the Second International Symposium on Maps and Graphics for Visually Impaired People Nottingham, UK, 1988.

Tehnologii de asistare a persoanelor cu deficiențe vizuale

- Tatham, Andrew F., *The design of tactile maps: theoretical and practical considerations*, Paper presented to the International Cartographic Association: mapping the nations, London, 1991.
- Thinus-Blanc, Catherine & Florence Gaunet, *Representation of space in blind persons: vision as a spatial sense?*, Psychol Bull 121.20-42, 1997.
- Timpf, Sabine, Gary Volta, David Pollock & Max Egenhofer, *A conceptual model of wayfinding using multiple levels of abstraction*, Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, ed. by A. Frank, I. Campari & U. Formentini, 348-67: Springer Berlin / Heidelberg, 1992.
- Tinti, Carla, Mauro Adenzato, Marco Tamietto & Cesare Cornoldi, *Visual experience is not necessary for efficient survey spatial cognition: evidence from blindness*, Q.J.Exp.Psychol (Colchester.) 59.1306-28, 2006.
- Tolman, Edward C., *Cognitive maps in rats and men*, Psychological review 55.189, 1948.
- Ungar, Simon, Mark Blades & Christopher Spencer, *The role of tactile maps in mobility training*, British Journal of Visual Impairment 11.59, 1993.
- Ungar, Simon, Mark Blades & Christopher Spencer, *Visually impaired children's strategies for memorising a map*, British Journal of Visual Impairment 13.27-32, 1995.
- Velázquez, Ramiro & Edwige E. Pissaloux, *Tactile displays in human-machine interaction: four case studies*, International Journal of Virtual Reality 7.51-58, 2008.
- Wang, Zheshen, Baoxin Li, Terri Hedgpeth & Teresa Haven, *Instant tactile-audio map: enabling access to digital maps for people with visual impairment*, Paper presented at the ASSETS'09, Pittsburgh, Pennsylvania, USA, 2009.
- White, Benjamin, Frank A. Saunders, Lawrence Scadden, Paul Bach-Y-Rita & Carter C. Collins, *Seeing with the skin*, Attention, Perception, & Psychophysics 7.23-27, 1970.
- Wijntjes, Maarten W., Thijs van Lienen, Ilse M. Verstijnen & Astrid M. Kappers, *The influence of picture size on recognition and exploratory behaviour in raised-line drawings*, Perception 37.602-14, 2008.
- Wolff, Marion, Philippe Cabon, Gérard Uzan, Julien Nelson & Stanislas Couix, *Déplacement urbain de personnes non-voyantes: étude multi-factorielle des difficultés et apport d'une nouvelle interface pour le recueil des données*, Paper presented at the Ergo-IA 2006, Biarritz, 2006.
- Wright, Tessa, Beth Harris & Eric Sticklen, *A Best-Evidence Synthesis of Research on Orientation and Mobility Involving Tactile Maps and Models*, JVIB 104.95-106, 2010.
- Zwiers, Marcel P., John A. Van Opstal & Johan R. Cruysberg, *A spatial hearing deficit in early-blind humans*, Journal of Neuroscience 21.RC142-RC45, 2001a.
- Zwiers, Marcel P., John A. Van Opstal & Johan R. Cruysberg, *Two-dimensional sound-localization behavior of early-blind humans*, Experimental Brain Research 140.206-22, 2001b.
- Zwiers, Marcel P., John A. Van Opstal & Gary D. Paige, *Plasticity in human sound localization induced by compressed spatial vision*, Nature Neuroscience 6.175-81, 2003.

Dispozitive imagistice dezvoltate de Centrul de Excelență Secolul 21 – Grupul Universității Shizuoka O introducere în nanovision

HIDENORI MIMURA¹

ABSTRACT. The 21st Center of Excellence (COE) group in Shizuoka University has proposed the nanovision science as a novel imaging science. An aim of the nanovision science is to develop imaging devices having ultimate performance in temporal, spatial, dynamic or wavelength ranges by introducing nanotechnology into imaging. As nanovision devices, we describe a time-resolved CMOS image sensor for time-domain fluorescence lifetime imaging, using an electron beam excitation assisted-optical microscope and a CdTe X-ray image sensor, capable of identifying the atomic number and electron density.

KEYWORDS: *nanovision science, time-resolved CMOS image sensor, time-domain fluorescence lifetime imaging, electron beam excitation assisted optical microscope, imaging of biological specimens, photon counting type CdTe X-ray image sensor, dual-energy X-ray CT, identifying atomic number and electron density.*

1. Introducere

Pentru a cunoaște lumea înconjurătoare, omul folosește cele cinci simțuri, dar cel puțin 80% din informații sunt achiziționate cu ajutorul *vederii*. Televiziunea, fiind bazată pe imagine, a promovat dezvoltarea democratică prin transmiterea în timp real, a unor evenimente care au loc oriunde în lume. Ea a condus la dezvoltarea economiei de consum printr-o publicitate puternică și influentă. Cu alte cuvinte, tehnologia bazată pe imagine a contribuit la crearea civilizației industriale moderne, având un impact incomparabil mai mare decât informația scrisă sau audio. Informația video a fost îmbogățită cu ajutorul tehnologiei informatice care a devenit indispensabilă în toate domeniile legate de informație, comunicare, divertisment, știință, artă, medicină, securitate.

¹ Profesorul Hidenori Mimura, Director Research Institute of Electronics, Shizuoka University, 3-5-1 Johoku Naka-ku Hamamatsu 432-8011, Japan, autor a acestui capitol, colaborare dezvoltată ca urmare a participării la conferința InterAcademia 2012, Budapesta, în cadrul proiectului „Societatea Bazată pe Cunoaștere – cercetări, dezbateri, perspective”, Uniunea Europeană și Guvernul României, Fondul Social European prin Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007-2013, POSDRU 56815, Academia Română: capitol tradus, editat, adaptat, de Mihaela Costin.

Recent, grupul de cercetare 21st Century of Excellence (CE) din cadrul Universității Shizuoka a propus utilizarea imagisticii cu rezoluție nanometrică (NV-nanovision)². Unul dintre obiectivele NV constă în dezvoltarea unor dispozitive de imagistică cu performanțe de ultimă generație, în domeniul temporal, spațial, dinamic sau al lungimii de undă prin utilizarea nanotehnologiei.

Capitolul de față descrie următoarele dispozitive utilizate în NV:

- un senzor de imagine TR-CMOS (TR – „time-resolved”, CMOS – abrevierea de la „complementary metal-oxide-semiconductor”), deci CMOS cu o bună rezoluție temporală, bazat pe măsurarea timpului de viață al fluorescenței³;
- un microscop optic pentru vizualizarea excitării cu un fascicul de electroni;
- un senzor de imagine din CdTe în domeniul razelor X⁴ care este capabil să identifice numărul atomic și densitatea de electroni⁵.

2. Senzor de imagine de tip TR-CMOS pentru imagistica timpului de viață al fluorescenței⁶

Imagistica de fluorescență este un instrument puternic utilizat în biologie. Fluorescența are două mărimi fizice caracteristice: intensitatea și durata.

Măsurarea intensității este dificilă, deoarece aceasta este influențată de concentrația de fluorofor, starea de degradare a acestuia, lungimea de undă și intensitatea fluxului de excitație, sensibilitatea detectorului și transmitanța sistemului optic. Pe de altă parte, măsurarea duratei fluorescenței utilizează rata de scădere a intensității. Astfel, cuantificarea, nu este influențată de degradarea intensității fluorescenței.

Metoda de numărare a unui singur foton, corelată în timp, este o metodă tipică folosită în microscopia imagistică a timpului de viață al fluorescenței (FLIM - fluorescence lifetime imaging microscopy). Senzorul TR-CMOS poate fi implementat relativ ieftin, într-un sistem FLIM compact. Figurile 1. (a) și (b) prezintă o diagramă bloc a măsurătorilor timpului de viață a fluorescenței și respectiv dependența numărului de fotoelectroni captați în fereastra de timp ca o funcție de t_{ds} în senzorii CMOS. Atunci când un impuls foarte scurt de lumină (de obicei mai mic de 100 ps) iradiază o probă fluorescentă, proba emite o fluorescență cu o scădere exponențială rapidă, de obicei între o nanosecundă și câteva zeci de nanosecunde.

² Hidenori Mimura, Kazuhiko Hara, Shoji Kawahito, Toru Aoki, and Norihisa Hiromoto, Nanovision Science-Evolution of Imaging Technology, Corona Pub. Co. Ltd., (2009) in Japanese.

³ Zhuo Li, Shoji Kawahito, Keita Yasutomi, Keiichi Kagawa, Juichi Ukon, and Hirohiko Niioka, A Time-Resolved CMOS Image Sensor With Draining-Only Modulation Pixels for Fluorescence Lifetime Imaging, IEEE Trans. Electron Devices, 59 (2012) 2715.

⁴ Wataru Inami, Kentaro Nakajima, Atsuo Miyakawa, and Yoshimasa Kawata, Electron beam excitation assisted optical microscope with ultra-high resolution, Optics Express 18 (2010) 12897.

⁵ Wenjuan Zou, Takuya Nakashima, Yoshiaki Onishi, Akifumi Koike, Bunji Shinomiya, Hisashi Morii, Yoichiro Neo, Hidenori Mimura, and Toru Aoki, Atomic Number and Electron Density Measurement Using a Conventional X-ray Tube and a CdTe Detector, Jpn. J. Appl. Phys. 47 (2008) 7317.

⁶ Hidenori Mimura, 2009, *Idem*.

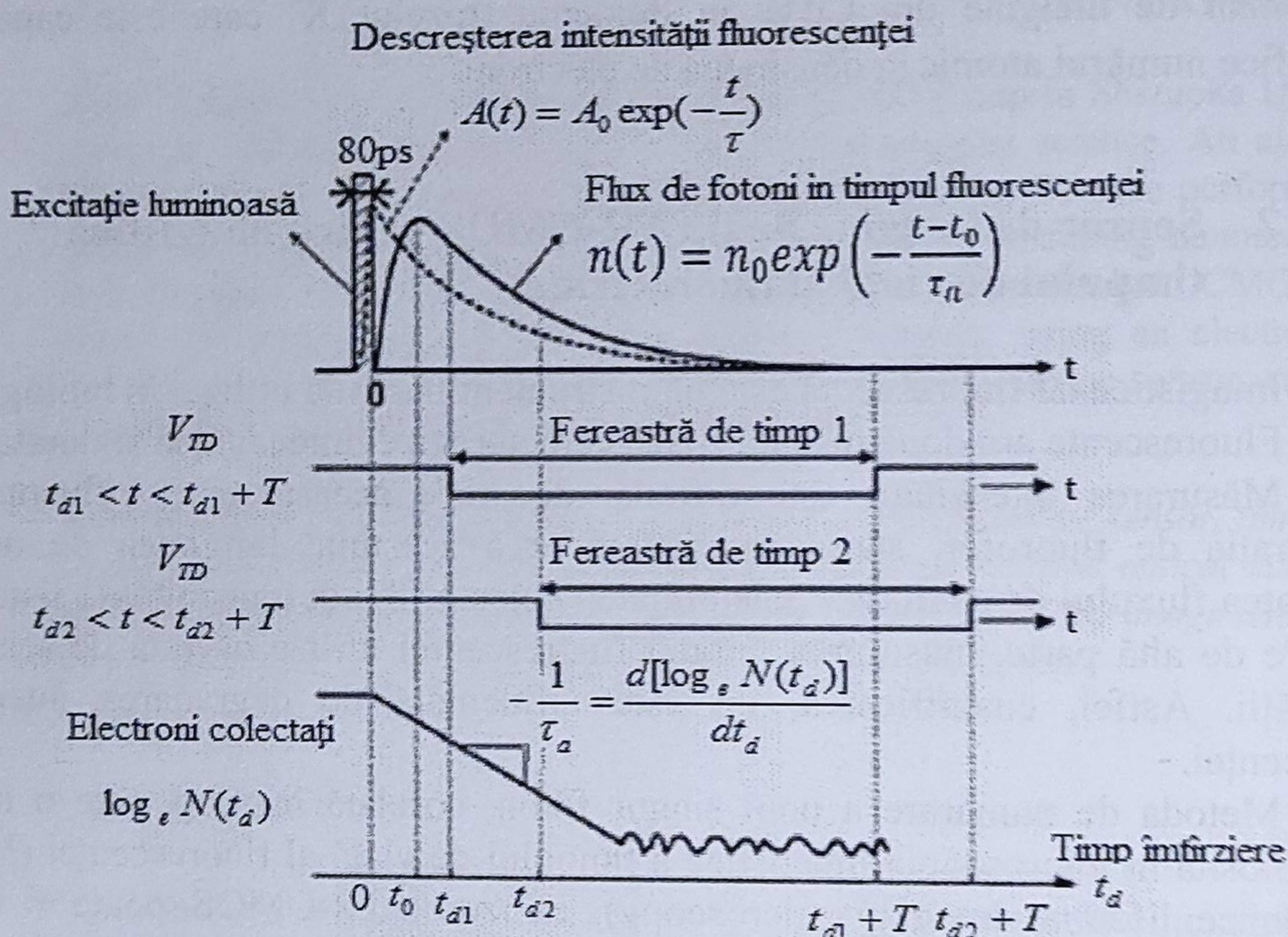
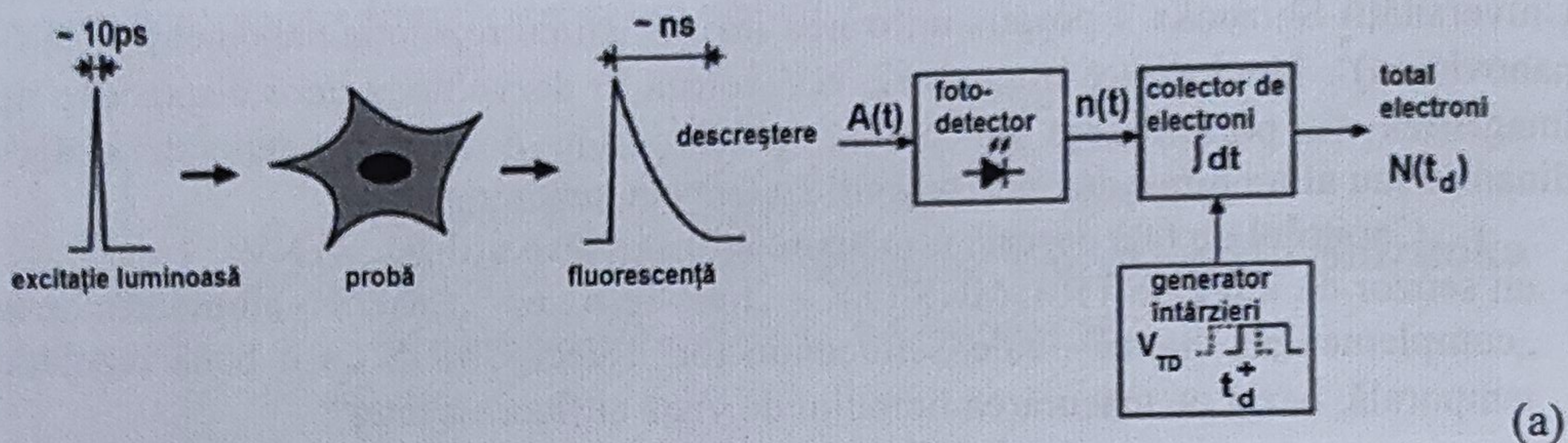


Fig. 1. (a) Diagramă bloc a măsurării timpului de viață a fluorescenței
(b) dependența numărului de fotoelectroni captați în fereastra de timp ca funcție de t_d în dispozitivul pentru generare de imagini, CMOS

Radiația de fluorescență incidentă este captată de un fotodetector care generează fotoelectroni. Numărul de electroni generați pe unitatea de timp este exprimat ca:

$$n(t) = n_0 \exp\left(-\frac{t-t_0}{\tau_a}\right) \quad (1)$$

pentru $t \geq 0$, unde n_0 este numărul inițial de electroni, $t-t_0$ este întârzierea generării fotoelectronului, iar τ_a este timpul de viață măsurat. Pentru măsurarea timpului de viață a fluorescenței, o parte a semnalului electronic de fluorescență este detectat și colectat într-o fereastră de timp. Fereastra de timp folosită la detecție este setată în funcție de tensiunea de poartă V_{TD} a fotodiodei.

Fotoelectronii generați sunt acumulați de un colector electronic în fiecare perioada T a ferestrei de timp. Punctul de start al ferestrei de timp este întârziat folosind un generator de întârziere, iar electronii sunt colectați de la momentul t_d până la momentul $t_d + T$.

Dacă t_0 este ales cu valoarea zero pentru a simplifica, numărul de electroni colectați $N(t_d)$ este exprimat ca:

$$\begin{aligned} N(t_d) &= \int_{t_d}^{t_d+T} n(t) dt = n_0 \tau_a \exp\left(-\frac{t_d}{\tau_a}\right) \text{ pentru } (t_d \geq 0) \\ N(t_d) &= \int_{t_d}^{t_d+T} n(t) dt = n_0 \tau_a \text{ pentru } (t_d < 0) \end{aligned} \quad (2)$$

Dacă T este ales suficient de mare, valoarea $\exp\left(-\frac{t_d + T}{\tau_a}\right)$ este neglijabilă.

Deoarece fluorescența unei probe biologice reale este foarte slabă, numărul mediu de fotoelectroni colectați pentru fiecare impuls de excitare este adesea mai mic decât 1. Pentru intensificarea semnalului, excitarea este repetată de multe ori, iar fotoelectronii sunt colectați în mod repetat în fiecare fereastră de timp (de aceeași durată). Numărul de electroni colectați descrește exponențial pentru început, urmat de un palier lung care este cauzat de fluctuațiile datorate zgomotului. Durata fluorescenței este obținută din panta descreșterii exponențiale a numărului de electroni colectați, care este dată de $-\frac{1}{\tau_a}$.

Dr. Yoon din grupul Prof. Kawahito a propus un senzor de imagine TR-CMOS pentru măsurarea timpului de viață a fluorescenței pe o durată de ordinul nanosecundelor⁷. Acest senzor are o problemă legată de sensibilitatea variației de la un pixel la altul în cazul unei iluminări foarte slabe. Pentru a rezolva această problemă, Dr. Li din același grup a propus un senzor de imagine TR-CMOS cu o structură a pixelului de tip DOM („draining-only modulation” – modulare prin drenare)⁸.

Figura 2 prezintă conceptul structurii DOM și profilul ei de potențial. O poartă de scurgere a sarcinilor (TD gate) este atașată pe lângă canalul purtătorilor de la PPD – fotodioda PIN (PPD – abreviere de la „pinned photodiode”, unde PIN - abrevierea de la „positive intrinsic negative diode”, iar PD – abrevierea de la „photodiode”) la PSD dioda de stocare (PSD – abrevierea de la „pinned storage diode”). La închiderea porții TD, sarcina este transferată către dioda PSD. La deschiderea porții, sarcinile din PPD sunt drenate înainte de a se deplasa către PSD. Operația de citire a semnalului de ieșire din PSD este realizată cu ajutorul porții de transfer (TX gate).

⁷ Hyung-June Yoon, Shinya Itoh, and Shoji Kawahito, A CMOS Image Sensor with in-Pixel Two-Stage Charge Transfer for Fluorescence Lifetime Imaging, IEEE Transactions on Electron Devices, Vol. 56, Nr. 2, (Febr. 2009) 214–221.

⁸ Zhuo Li, Shoji Kawahito, Keita Yasutomi, Keiichiro Kagawa, Juichiro Ukon, and Hirohiko Niioka, A Time-Resolved CMOS Image Sensor With Draining-Only Modulation Pixels for Fluorescence Lifetime Imaging, IEEE Trans. Electron Devices, 59 (2012) 2715.

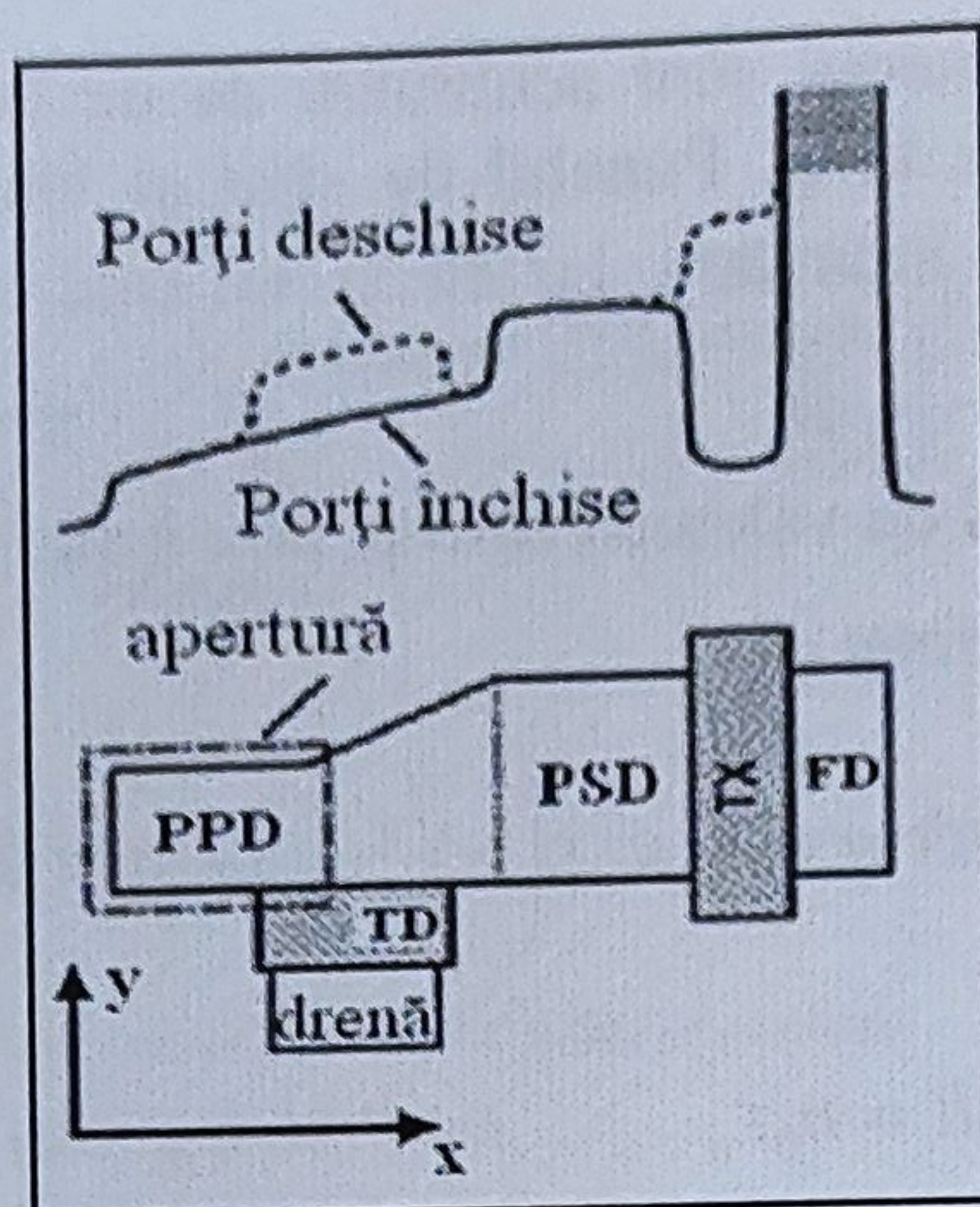


Fig. 2. Conceptul de realizare a structurii DOM și profilul potențialului

Figura 3 arată o imagine a întregului pixel DOM. Forma efectivă a PPD în display are câțiva pași care cresc gradual lărgimea canalului din zona de deschidere către PSD în loc de a avea o variație continuă în Fig 2. Pixelul DOS utilizează o tehnologie de imagine CMOS standard cu opțiune PPD. În condițiile unui câmp electric slab într-o diodă PIN săracită, transferul purtătorilor de-a lungul a câtorva micrometri poate avea loc în câteva nanosecunde sau mai puțin decât 1 ns. Acest timp de răspuns sub 1 ns este suficient pentru observarea fenomenelor rezolvate în timp cu o rezoluție mai mică decât 1 ns. Pixelul efectiv constă dintr-un PPD, un PSD, o poarta TD, o poarta TX între PSD și FD (FD – „floating diffusion”), difuzie flotantă, un transistor de reset, un repetor cu un comutator de selecție a pixelului. Partea principală a pixelului este simetrică de-a lungul liniei A-A', care conține două căi de transfer de sarcină, și poarta TD este situată lângă fiecare cale de transfer. Această structură simetrică permite ca aria fotosensibilă să fie crescută pentru creșterea sensibilității.

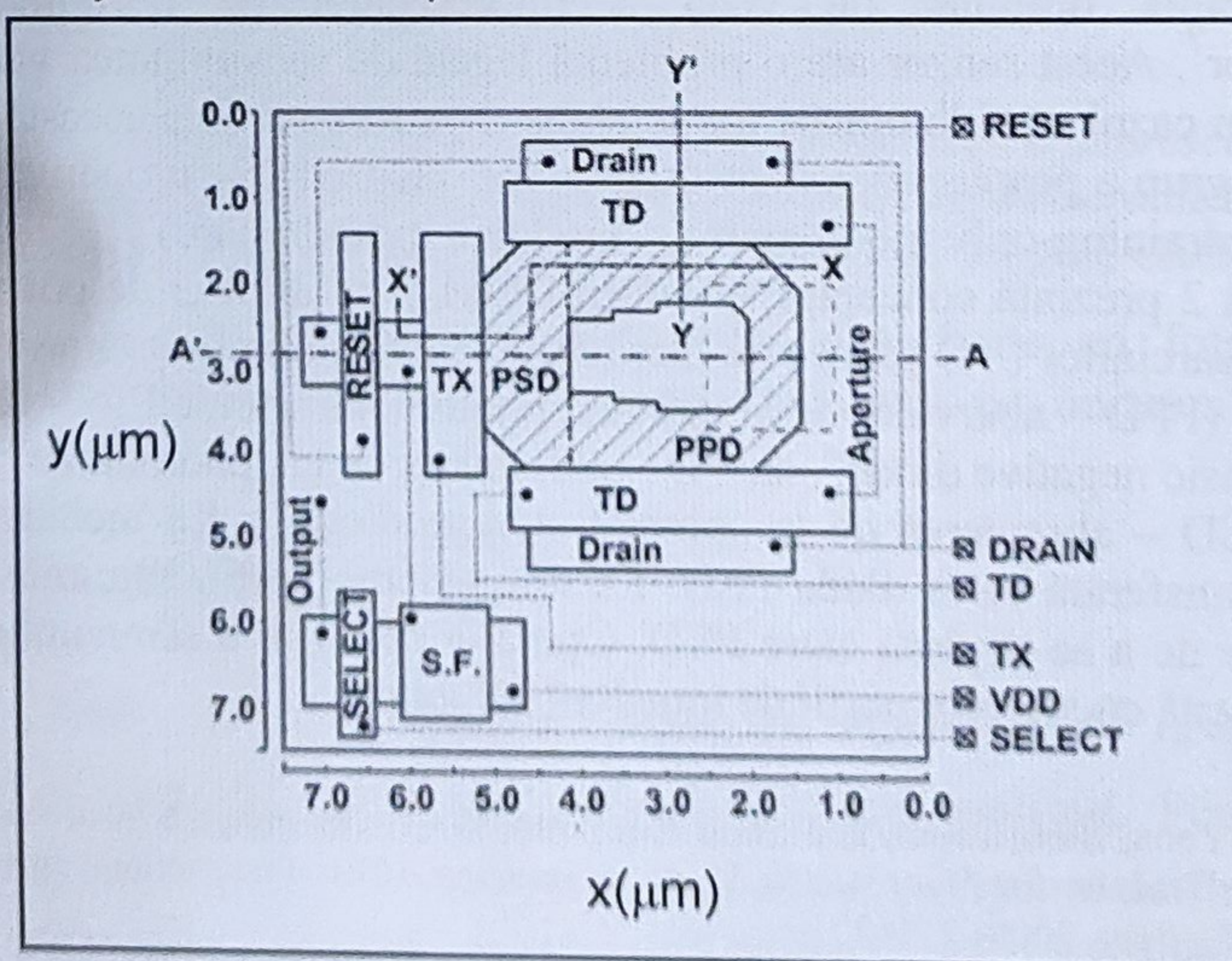


Fig. 3. Prezentarea structurii întregului pixel DOM

Figura 4 (a) and (b) reprezintă o secțiune transversală a pixelului și profilele de potențial de-a lungul direcțiilor: PPD – PSD (X-X') (a) și respectiv PPD – drenă (Y-Y') (b), pentru a explica mecanismul de modulare în sarcină a pixelului DOM.

Când poarta TD este închisă, se creează un câmp electric lateral pozitiv monoton pentru transfer de sarcină la viteză înaltă din PPD la PSD așa cum se observă în curba punctată din figura 4 (c) care arată funcționarea pentru modularea sarcinii foto generate. Când poarta TD este deschisă, o scădere de potențial se produce în calea de transfer a purtătorilor, așa cum se observă în curba solidă (continuă) din fig. 4d. Purtătorii care cad în această groapă de potențial sunt drenați prin poarta TD. Operația de citire a electronilor generați de fluorescent este ilustrată în fig. 4e. După detectarea repetată a semnalului din fereastră, semnalul stocat în PSD este citit utilizând o operație de dublă eșantionare corelată (CDS - correlated double sampling). Pentru aceasta, mai întâi se resetează FD, și apoi sarcina din PSD este transferată la FD prin procesul de citire, în timp ce TD este întotdeauna deschisă pentru drenarea electronilor de întuneric în PPD.

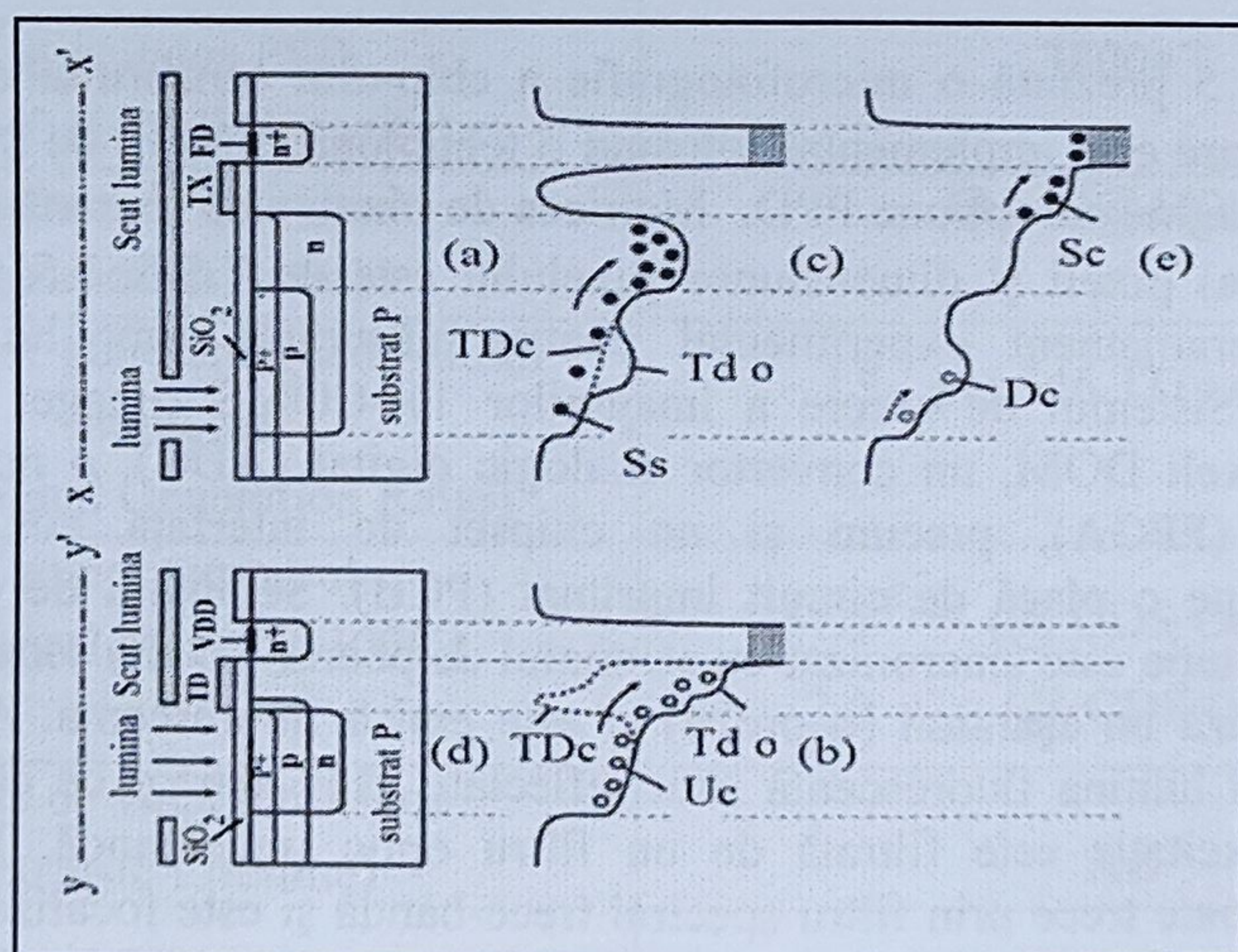


Fig. 4. Secțiune transversală a „pixelului” și profilele de potențial de-a lungul direcțiilor: PPD – PSD (X-X') (a) și PPD - drenă (Y-Y') (b)(c) și (d) arată funcționarea pentru modularea sarcinii foto generate (e) arată semnalul de citire secvențială în pixelul DOM

Un semnal de fluorescent dintr-o fereastră de timp, poate fi detectat astfel: în timpul excitației cu impulsul de lumină, poarta TD este deschisă pentru a drena sarcinile nedorite generate direct de pulsul de lumina incidentă. Ca urmare se emite radiație de fluorescent care descrește exponential și care generează sarcini-semnal. Poarta TD rămâne deschisă până la începutul ferestrei de timp. În timpul ferestrei de timp, de la t_{d1} la $t_{d1}+T$, poarta TD este închisă. Ca urmare, o parte din semnalul descrescător generat în PPD este transferat în PSD. Procesul este repetat de multe ori utilizând ferestre de timp cu aceeași întârziere față de pulsul de lumină excitatoare pentru a intensifica slabul semnal de fluorescență.

Pentru a detecta electronii semnal într-o altă fereastră de timp, se utilizează o altă întârziere temporală t_{d2} arătată în Fig. 1a și 1b.

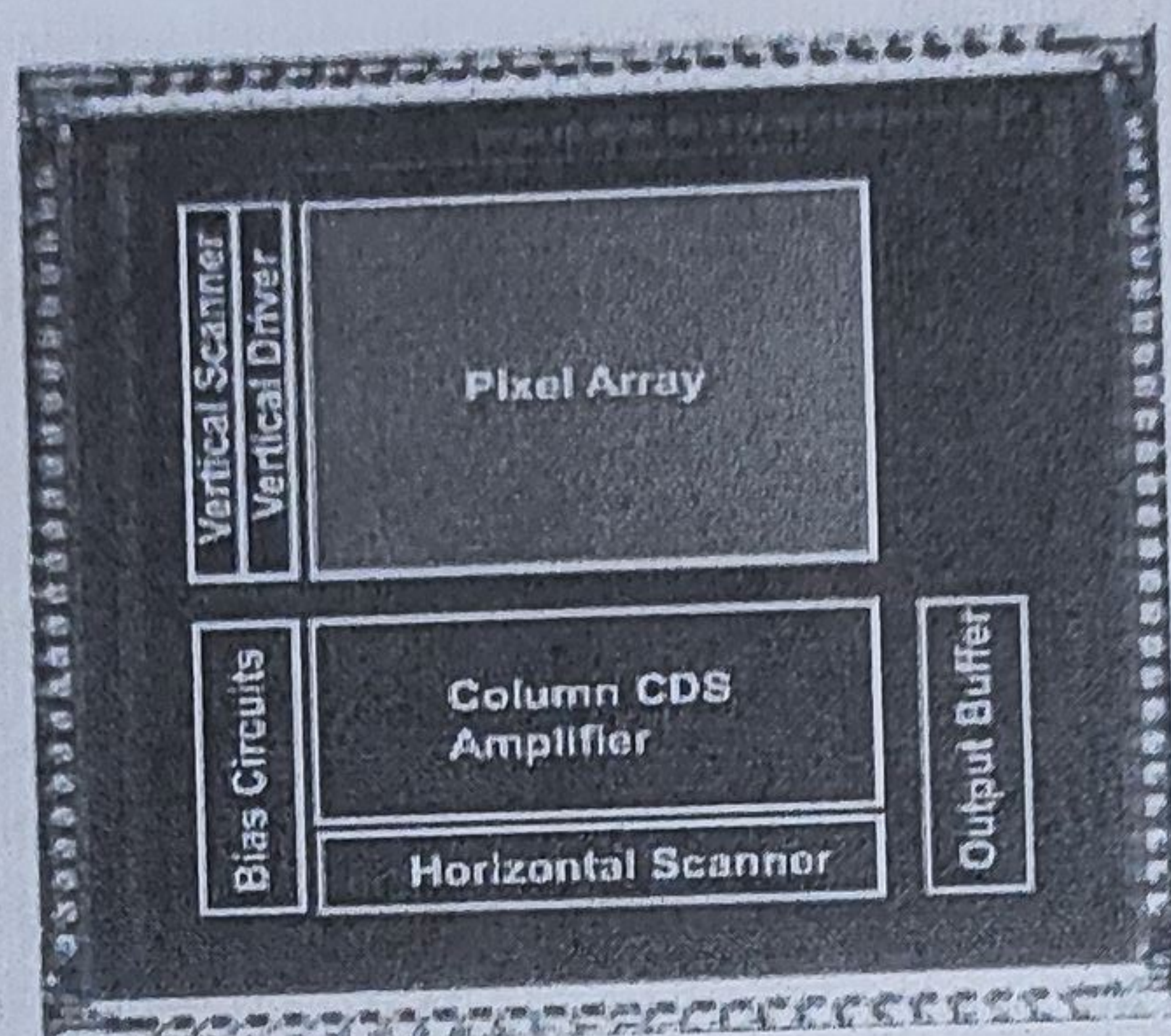


Fig. 5. Microfotografie a chip-ului senzorului de imagine cu pixeli DOM

Figura 5 prezintă o microfotografie a chip-ului senzorului de imagine cu pixeli DOM, care este implementat folosind o tehnologie CMOS de $0.18\ \mu\text{m}$ pentru senzorul de imagine cu opțiune PPD. Matricea de pixeli este formată din 256 (rând) x 256 (coloana) pixeli și dimensiunea pixelului este de $7.5 \times 7.5\ \mu\text{m}^2$. Figura 6 prezintă un aranjament experimental pentru măsurarea timpului de viață al fluorescenței. Sistemul de creare a imaginilor TR-COMS (imager time-resolved CMOS) cu pixeli DOM, un convertor analogic digital (ADC), o matrice de porți programabile (FPGA), precum și un chipset de interfață cu camera, sunt implementate pe o placă de circuit imprimat (PCB). Semnalul de control pentru lumina de excitație este sincronizat cu semnalul de poartă TD din senzor. Lumina de excitație iradiază un eșantion (o probă) și este emisă fluorescența. Atât lumina de excitație cât și lumina fluorescentă sunt reflectate către senzor de către o oglindă. Lumina de excitație este filtrată de un filtru optic trece-bandă. Numai lumina fluorescentă poate trece prin filtru spectral trece-bandă și este focalizată pe senzorul de imagine printr-o lentilă. Senzorul de imagine detectează lumina fluorescentă. Fereastra timpului de detecție (v. Fig. 1), este setată la 125 ns. Ieșirea semnalului analogic al senzorului este conectată la un ADC („analog-to-digital converter”) pentru fi convertită într-un cod digital de 14b. Un FPGA este folosit pentru generarea semnalelor digitale, pentru controlul senzorului și recepționarea codului digital de la ADC. Un protocol de conexiune cu camera (camera-link) este folosit pentru a transmite codul digital de la FPGA la un calculator. După prelucrarea datelor, imaginea este afișată pe un monitor, iar datele sunt stocate în calculator. Tabelul 1 prezintă sinteza performanței prototipului chip-ului senzorului.

Parametri	Valoare
Pixeli efectivi (Effective Pixels)	256 (Coloana) x 256 (Rand)
Mărimea pixelului (Pixel size)	$7.5\mu\text{m} \times 7.5\mu\text{m}$
Rata frame-urilor, cadrelor (Frame Rate)	15fps
Factorul de umplere (fără micro-lentile) (Fill Factor - without microlens)	4.6 %
Amplificarea la conversie (Conversion Gain)	$30\mu\text{V}/e^-$
Capacitatea PSD (PSD Full Well Capacity)	$3800e^-$
Raportul de extincție) (Extinction Ratio)	30 : 1 (@440nm)
Raportul de generare a semnalului întunecat (Dark Signal Generation Ratio)	$1 e^-$ per 546 TD impulsuri
Zgomot temporal (Temporal Noise @ 300K)	$17.2 e_{rms}^-$ (@ CDS Gain 1.07 x) $2.0 e_{rms}^-$ (@ CDS Gain 30 x)
Durata de viață intrinsecă (Intrinsic Lifetime)	2ns

Tabelul 1. Prezentarea performanțelor prototipului chip-ului senzor

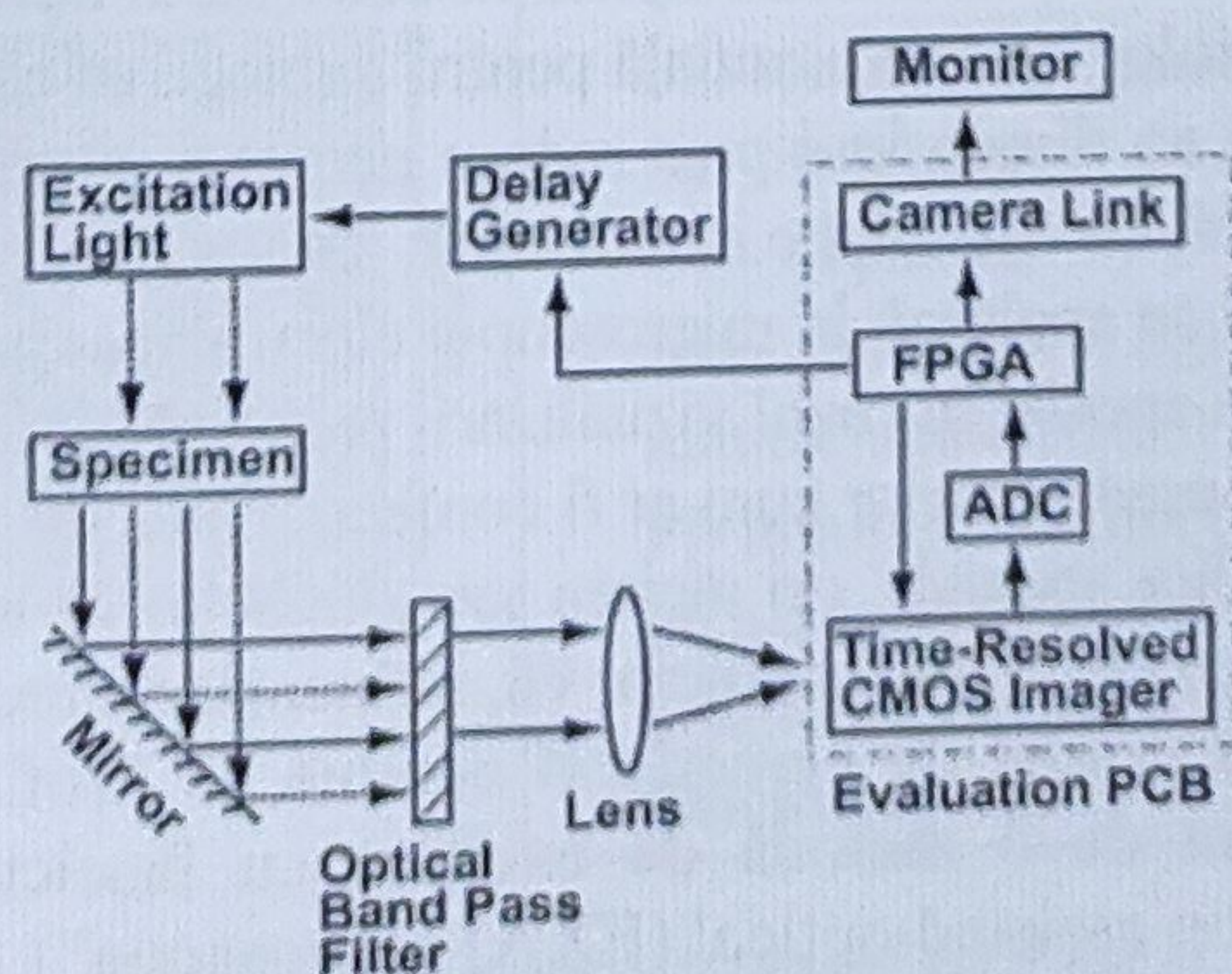


Fig. 6. Structura experimentală a măsurătorilor pentru timpul de viață

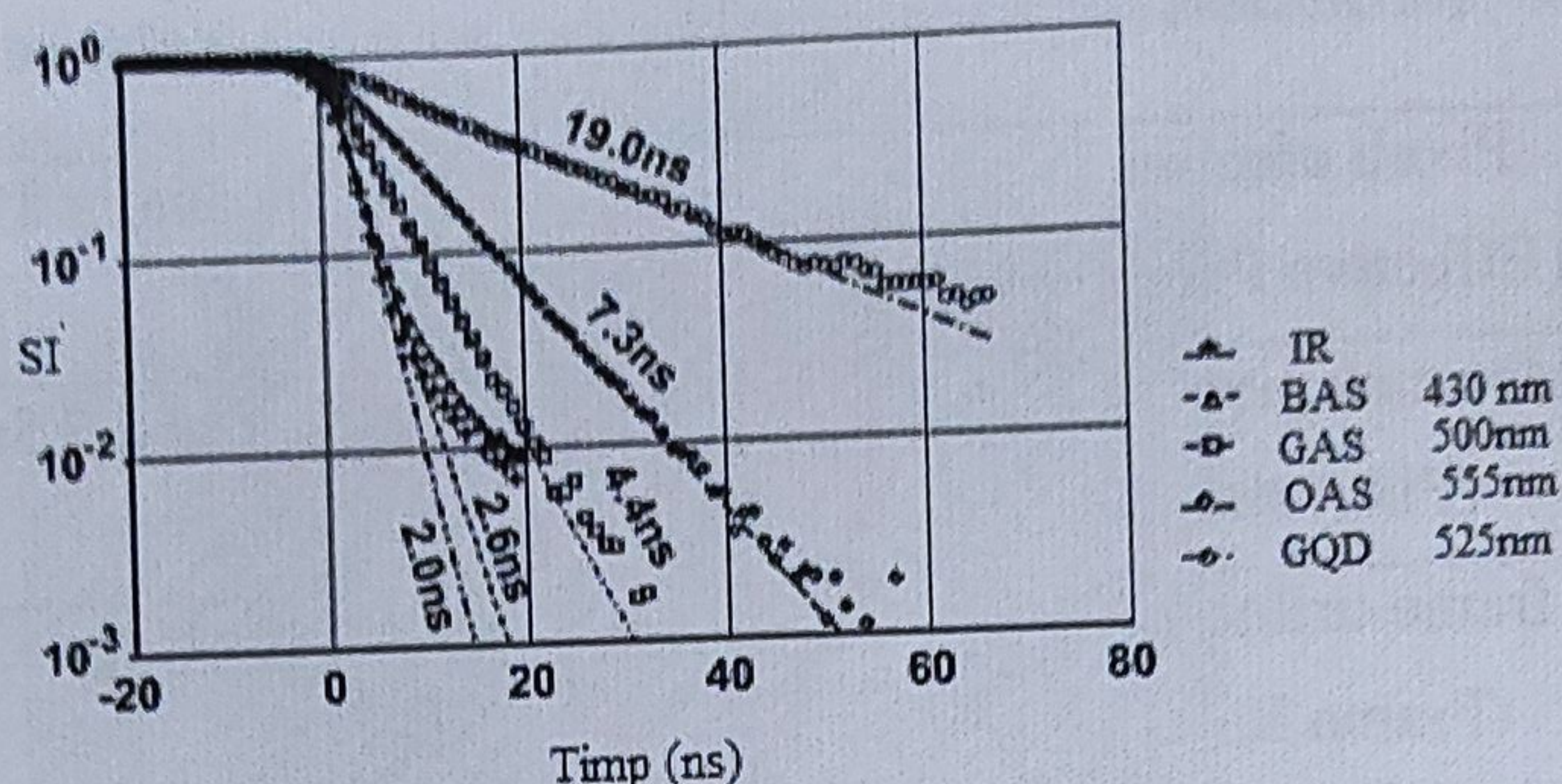


Fig.7. Măsurarea timpului de viață pentru răspunsul intrinsec al senzorului (curbele de descreștere a intensității semnalului pentru emisia fluorescentă) - patru tipuri de fluorofori.

Figura 7 reprezintă curbele de descreștere a intensității semnalului pentru emisia luminii de fluorescență pentru patru tipuri de fluorofori și răspunsul intrinsec al acestui sistem de imagistică rezolvată temporal (time-resolved imager). Abrevierile sunt: SI-Signal Intensity (Normalized) – Intesitatea semnalului (normalizat); IR-Intrinsic Response răspuns intrinsec; BAS-Blue Acrylic Screen, ecran acrilic albastru; GAS-Green Acrylic Screen - ecran acrilic verde; OAS-Orange Acrylic Screen, ecran acrilic portocaliu; GQD-Green Quantum Dot. Se utilizează pentru toate măsurătorile o diodă laser cu radiație ultravioletă cu o lungime de undă de 374 nm și un puls de 80ps. Pixelul DOM cu un timp de transfer de sarcină finit, are timpul de viață intrinsec τ_0 , care este determinat de dispersia spotului de lumină pe fotodiodă și de dispersia timpului de transfer. În măsurarea timpului de viață intrinsec, lumina de excitare este direcționată pe chip-ul senzorului.

3. Microscopul optic asistat cu excitare cu fascicul de electroni

Achiziționarea de imagini pentru eșantioane (probe) în domeniul nanometric de analizat (NV) este necesară în numeroase aplicații. Procesarea acestor imagini de mare rezoluție ale unor probe biologice este esențială pentru înțelegerea cât mai exactă a funcțiilor celulare. Multe structuri au dimensiuni prea reduse pentru a putea fi studiate folosind microscopie optice standard, cu o rezoluție insuficientă, datorată limitei de difracție a luminii.

În aceste cazuri se apelează la microscopia electronică de mare rezoluție, la nivel de nanometru. Totuși, în special în cazul vizualizării probelor biologice, sunt necesare preparative preliminare ale eșantioanelor cum ar fi confectionarea de eșantioane foarte subțiri, acoperirea cu folii metalice speciale, ori răcirea sau „înghețarea” eșantioanelor. Asemenea operații pregătitoare afectează studiul celulelor vii, influențând starea lor naturală.

Dr. Inami, din grupul de cercetători condus de Prof. Kawata, a propus un microscop optic asistat de o metodă de excitare cu fascicul de electroni – „the electron beam excitation assisted optical (EXA) microscope”).

Figura 8(a) ilustrează schema EXA propusă de Dr. Inami. Un fascicul de electroni este focalizat pe un film luminescent (nitrură de siliciu SiN). Eșantionul este plasat apoi direct pe peliculă. În figură sunt arătate imaginea mărită a filmului și a eșantionului. Sursa de iluminare nanometrică este excitată în filmul luminescent de către fasciculul de electroni focalizat. Sursa de lumină nanometrică iluminează eșantionul iar lumina împrăștiată sau transmisă este detectată de un fotomultiplicator (PMT). Pentru a obține un contrast corespunzător al imaginii, sursa de lumină este scanată prin scanarea cu fasciculul de electroni focalizat.

O structură de suport de eșantion este ilustrată în Fig. 8(b). O fantă pătrată de $100 \times 100 \mu\text{m}^2$ a fost confecționată pe substratul de siliciu și fixată (acoperită) cu un film SiN de grosime 50 nm. Focalizând fasciculul de electroni se limitează mărimea spotului optic ceea ce înlocuiește confecționarea fantelor de deschidere mică, cum este cazul în microscopia optică tradițională în câmp apropiat (NSOM).

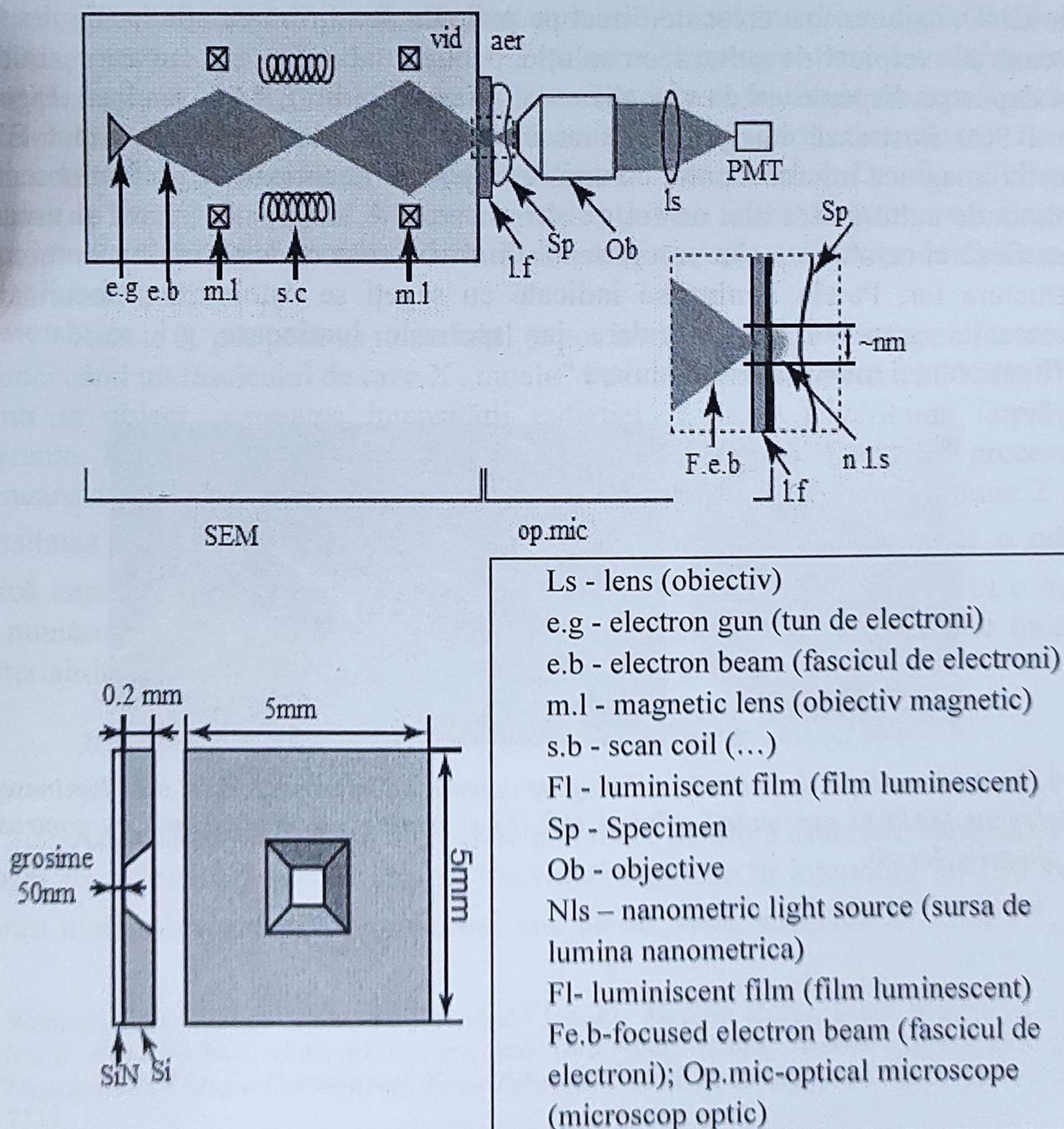


Fig. 8. Schema microscopului EXA (a) și (b) structura unei membrane pe bază de nitrură de siliciu

Un spot optic de câteva zeci de nanometri este produs ușor de un fascicul de electroni focalizat pe o suprafață de câțiva nanometri. Prin captarea imaginilor microscopice în secvențe de frecvență video sau mai mare, pot fi observate, activități în desfășurare ale eșantioanelor biologice vii, deoarece fasciculul de electroni poate fi scanat prin modularea câmpului electric sau magnetic fără a fi necesare componente mecanice în mișcare. Un raport semnal util/zgomot foarte bun se poate obține utilizând pelicule luminescente eficiente la iradierea cu fascicul. Un alt avantaj special al microscopului EXA constă în posibilitatea observării și captării imaginilor eșantioanelor aflate în diverse medii: aer, gaze, lichid sau vid. Filmul luminescent separă aerul sau lichidele din jurul eșantionului de vidul unde este focalizat fasciculul de electroni. Din moment ce energia fasciculului este convertită în lumină pe pelicula luminescentă, iar lumina transmisă și reflectată de eșantion este captată ca semnal, nu este necesară plasarea eșantioanelor în vid. Pentru a proba potențialul metodei propuse, în cazul celulelor vii, au fost vizualizate microorganisme MARCO, cu celule CHO, care au fost crescute direct pe pelicula de nitrură de siliciu. Figura 9(a) ilustrează un recipient de cultură, cu soluție, pentru studiu. După incubare eșantionul a fost depus pe dispozitivul de vizualizare al microscopului EXA și analizat. Figurile 9(b) și 9(c) ilustrează imagini de luminescență a celulelor cu microscopul EXA, respectiv imaginea microscopului cu contrast de fază. Eșantioanele au fost observate în soluția de cultură fără nici un fel de alte tratamente, cum ar fi fixarea și uscarea. Forma fiecărei celule este clar recunoscută fiind observate și unele spoturi luminoase în structura lor, Petele luminoase indicate cu săgeți se datorează probabil autofluorescenței granulelor intracelulare iar spoturile luminoase gri se datorează autofluorescenței membranelor celulare.

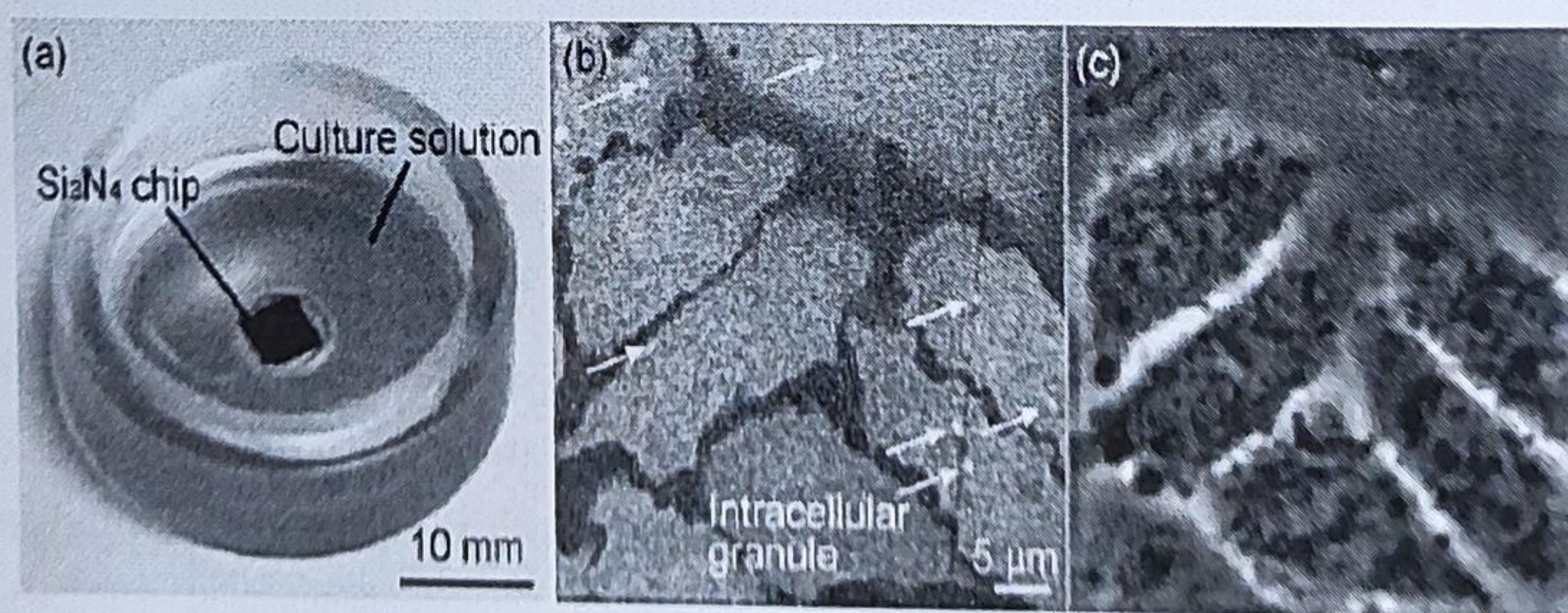


Fig. 9. Probă cu soluție de cultură (a), imagine captată cu microscop EXA prin luminescența celulelor vii MARCO exprimând celule CHO (b) și imagini de microscopie cu contrast de fază ale celulelor vii.

4. Senzor de imagine CdTe cu raze X, pentru identificarea numărului atomic și a densității de electroni⁹

Sistemele de detecție a explozibililor sunt importante în domeniul comunicațiilor și al transportului.

Până în prezent, s-a aplicat pe scară largă tehnica bazată pe dispozitive care funcționează cu un singur tip de energie. Dar pe baza acestei tehnici se pot efectua doar măsurători aproximative ale densității obiectelor scanate. Pentru identificarea materialului, o tehnică cu două tipuri de energie este o alternativă atractivă. Metoda tomografiei computerizate (CT) cu două energii, cu raze X, este o tehnică performantă pentru identificarea atât a numărului atomic cât și a densității de electroni a unui obiect scanat¹⁰, care aplică tehnica cu două energii. Pentru a folosi aceasta tehnică, trebuie să se cunoască valorile coeficienților de atenuare liniară la interacțiunea obiectului cu razele X incidente, corespunzător la două energii diferite. Experimental, acești coeficienți pot fi măsurați numai în cazul în care radiația incidentă și radiația detectată, sunt monocromatice și au aceeași energie. Pentru a obține această condiție experimentală „ideală”, unele grupuri au folosit fascicule de raze X emise de accelerator cu putere foarte mare care să penetreze prin monocromatoare. Pentru a aplica tehnica cu energie dublă la identificarea materialului, Dr. Zou din grupul condus de Prof. Aoki și Prof. Mimura a propus un nou sistem de scanare tomografică computerizată folosind un tub de raze X convențional și un detector de CdTe pentru numărarea fotonilor¹¹.

Razele X pot penetra materialul și pot interacționa cu toate speciile atomice. Atunci când un fasciculul de raze X „moale” cu o energie mai mică de 200 KeV trece printr-un obiect, atenuarea intensității radiației X poate fi atribuită împrăstierii coerente, împrăstierii necoerente și absorbției fotoelectrice. Deoarece procesul de atenuare depinde de proprietățile materialului, și anume de numărul atomic Z și de densitatea de electroni ρ_e , coeficientul de atenuare liniară – care este o mărime fizică importantă pentru descrierea acestui proces – poate fi reprezentat ca o funcție de numărul atomic și de densitatea de electroni. Coeficientul de atenuare liniară al materialului pentru un fascicul de raze X cu energia E este scris ca:

$$\mu(E, Z, \rho_e) = \rho_e [\sigma^{coh}(E, Z) + \sigma^{incoh}(E, Z) + \sigma^{ph}(E, Z)] \quad (1)$$

Această formulă oferă o precizie mai mare pentru a descrie variația fiecărui proces de atenuare pentru o energie cu valori cuprinse în intervalul 30-150 keV și pentru materiale cu număr atomic mai mic de 50. Aici, termenii σ^{coh} , σ^{incoh} , σ^{ph}

⁹ Wenjuan Zou, Takuya Nakashima, Yoshiaki Onishi, Akifumi Koike, Bunji Shinomiya, Hisashi Morii, Yoichiro Neo, Hidenori Mimura, and Toru Aoki, *Atomic Number and Electron Density Measurement Using a Conventional X-ray Tube and a CdTe Detector*, Jpn. J. Appl. Phys. 47 (2008) 7317.

¹⁰ Yumiko Ohno, Masami Torikoshi, Takanori Tsunoo, and Kazuyuki Hyodo, *Dual-energy X-ray CT with CdTe Array and its extension*, Nuclear Instruments and Methods in Physics Research Section A548, pp 72-77, (2004)

¹¹ Wenjuan Zou, 2008, *Idem*.

sunt secțiunile eficace ale electronilor, determinate de împrăștierea coerentă, împrăștierea necoerentă și respectiv de absorbția fotoelectrică. Simplificată, această formulă este următoarea:

$$\mu(E, Z, \rho_e) = \rho_e [Z^4 F(E, Z) + G(E, Z)] \quad (2)$$

În ecuația (2), cei doi termeni din partea dreaptă, sunt determinați de efectul fotoelectric și de împrăștiere. Funcțiile $F(E, Z)$ și $G(E, Z)$ pot fi determinate prin compararea ec. (2) cu ec. (1). Din ec. (2), dacă se măsoară coeficienții de atenuare liniară a materialului pentru două valori distincte ale energiei, obținem:

$$Z^4 = \frac{\mu(E_2)G(E_1, Z) - \mu(E_1)G(E_2, Z)}{\mu(E_1)G(E_2, Z) - \mu(E_2)G(E_1, Z)} \quad (3)$$

Deși o expresie analitică pentru numărul atomic nu poate fi dedusă direct din această ecuație, soluția numerică pot fi obținută folosind un algoritm iterativ. Odată ce numărul atomic este cunoscut, densitatea de electroni poate fi, de asemenea, determinată din următoarea ecuație:

$$\rho_e = \frac{\mu(E_1)F(E_2, Z) - \mu(E_2)F(E_1, Z)}{F(E_2, Z)G(E_1, Z) - F(E_1, Z)G(E_2, Z)} \quad (4)$$

Dioda cu CdTe a fost fabricată folosind o metodă de dopare cu laser cu excimer¹². Figura 10 arată o instalație de dopare cu laser cu excimer. Un strat subțire de indiu cu o grosime de 30nm a fost depus pe o suprafață de CdTe (pe fața cu Te) într-un sistem convențional de evaporare în vid fără încălzirea substratului.

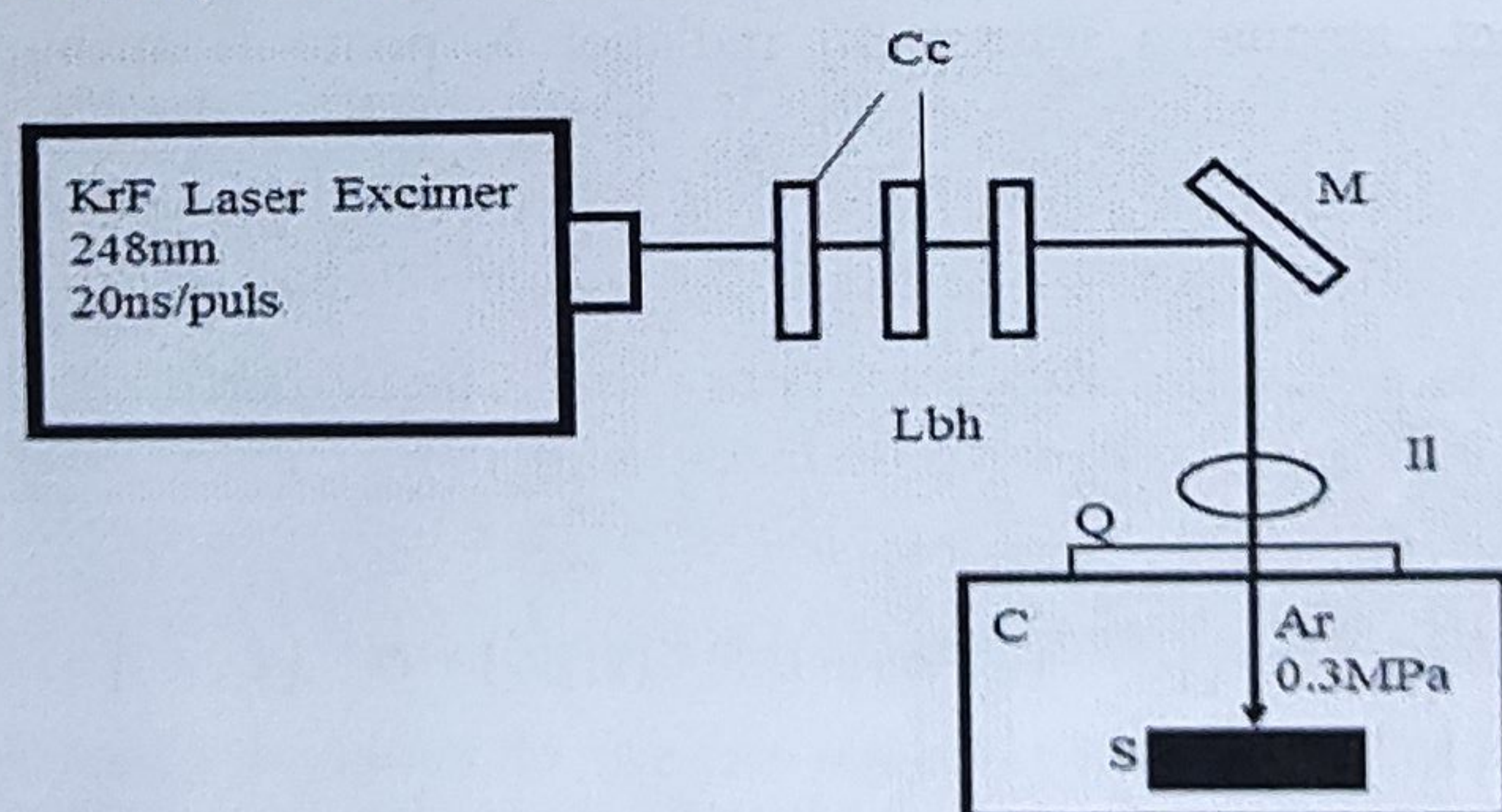


Fig.10. Instalație de dopare cu laser cu excimer. Cc Colimator cilindric; Lbh omogenizator pentru fascicul laser; lentile pentru imagine II; cuarț Q; cameră C; eșantion S; Oglindă M.

Am folosit un laser excimer KrF cu o lungime de undă de 248 nm, un impuls de lățime de 20nm și o putere de 75mJ/cm². Pulsul laser a fost introdus și direcționat pe suprafața de indiu a eșantionului într-o cameră de înaltă presiune cu o atmosferă

¹² Toru Aoki, Y. Ishida, D. Sakashita, V. Gnatyuk, A. Nakamura, Y. Tomita, Y. Hatanaka, and J. Tenmmyo, *Development of Energy-Discriminate CdTe Imaging Detector for Hard X-ray*, Proc. SPIE 5540 (2004) 196.

de 0.3 MPa cu argon, printr-o lentilă de cuarț. Un strat de CdTe de tip n, a fost obținut printr-un proces de difuzie termică. Apoi, s-a depus electrodul de Indiu de pe suprafața dopată și electrodul din aur a fost depus pe partea opusă. Figura 11 prezintă o caracteristică tipică curent-tensiune a unei diode CdTe la temperatura camerei. Mărimea dispozitivului e de 5mm x 5mm x 0.5mm. Curentul invers este mai mic decât câțiva nA pentru o tensiune aplicată de 100V.

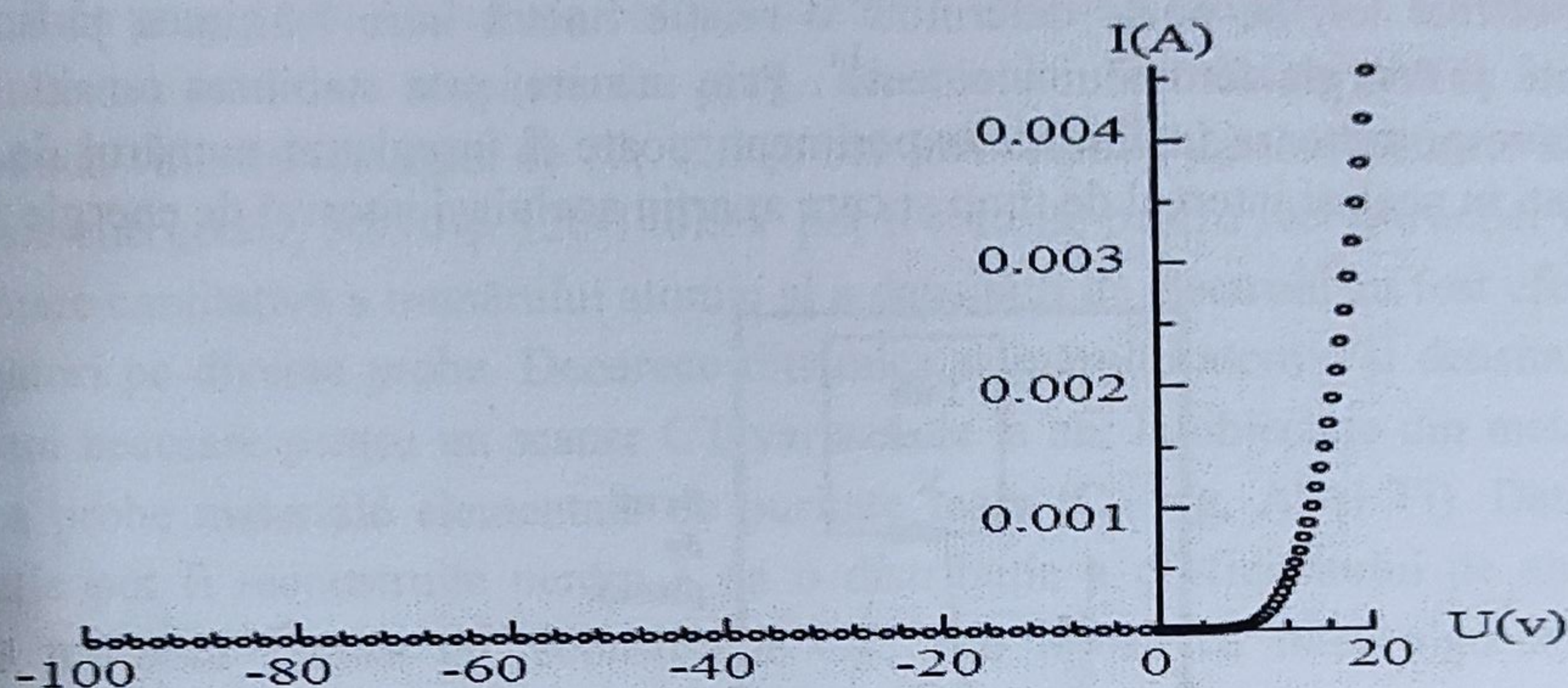


Fig. 11. Caracteristica curent-tensiune a diodei CdTe la temperatura camerei
I-curentul (A-Amper); U-tensiunea(V-volti)

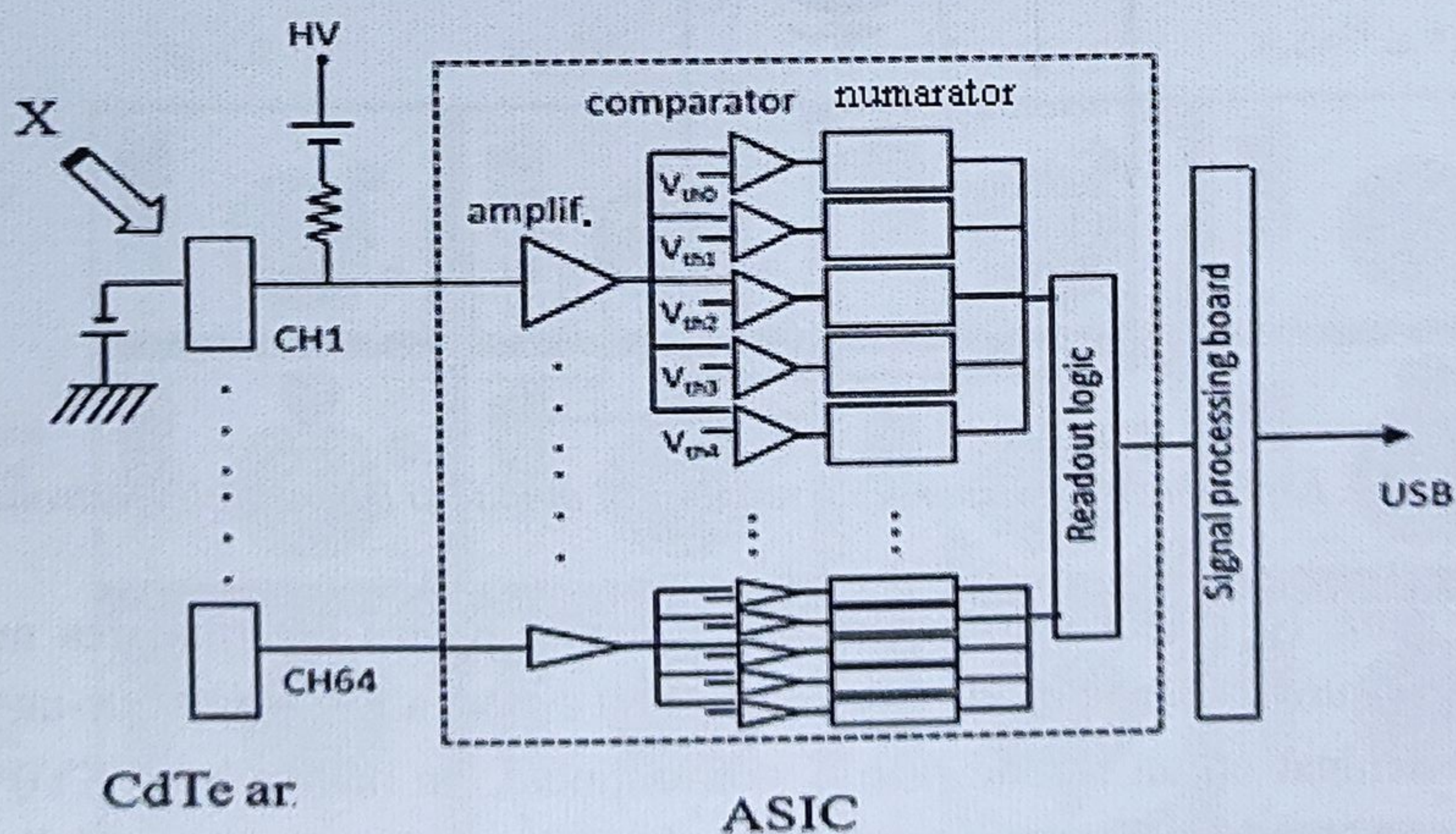


Fig.12. Diagrama detectorului CdTe - schema pentru numărarea fotonilor

Figura 12 prezintă o diagramă a unui detector de CdTe pentru numărarea fotonilor folosind o diodă CdTe. S-a folosit un senzor de imagine lineară CdTe cu raze X care este format din 64 de elemente cu dimensiunile: 0.8mm lățime, 0,5mm înălțime și 2mm adâncime¹³. Toate elementele din CdTe au fost aliniate una lângă alta, la distanță de 0,1 mm, și funcționează ca o matrice de detectare a radiațiilor. Odată ce un foton de radiație X este absorbit de către un element CdTe, este generată

¹³ Wenjuan Zou, 2008, *Idem*.

o sarcină electrică datorită efectului fotoelectric intern. Sarcina electrică generată este amplificată de către un amplificator și transformată într-un puls electric cu înălțimea proporțională cu energia fotonului incident. Pulsurile rezultate sunt comparate cu cinci tensiuni de prag, pe baza amplitudinilor lor și apoi sunt diferențiate și numărate de către contoarele relative. Prin măsurarea spectrului radiației a trei radioizotopi diferiți (^{57}Co , ^{241}Am , și ^{137}Cs), care au valori cunoscute ale energiei peak-lui principal din spectrele lor, se poate determina o relație liniară între înălțimea pulsului de tensiune și energia fotonului incident¹⁴. Prin urmare, prin stabilirea tensiunilor de prag corespunzătoare înainte de experiment, poate fi înregistrat numărul de fotoni detectați în același interval de timp și care aparțin aceluiași interval de energie fixat.

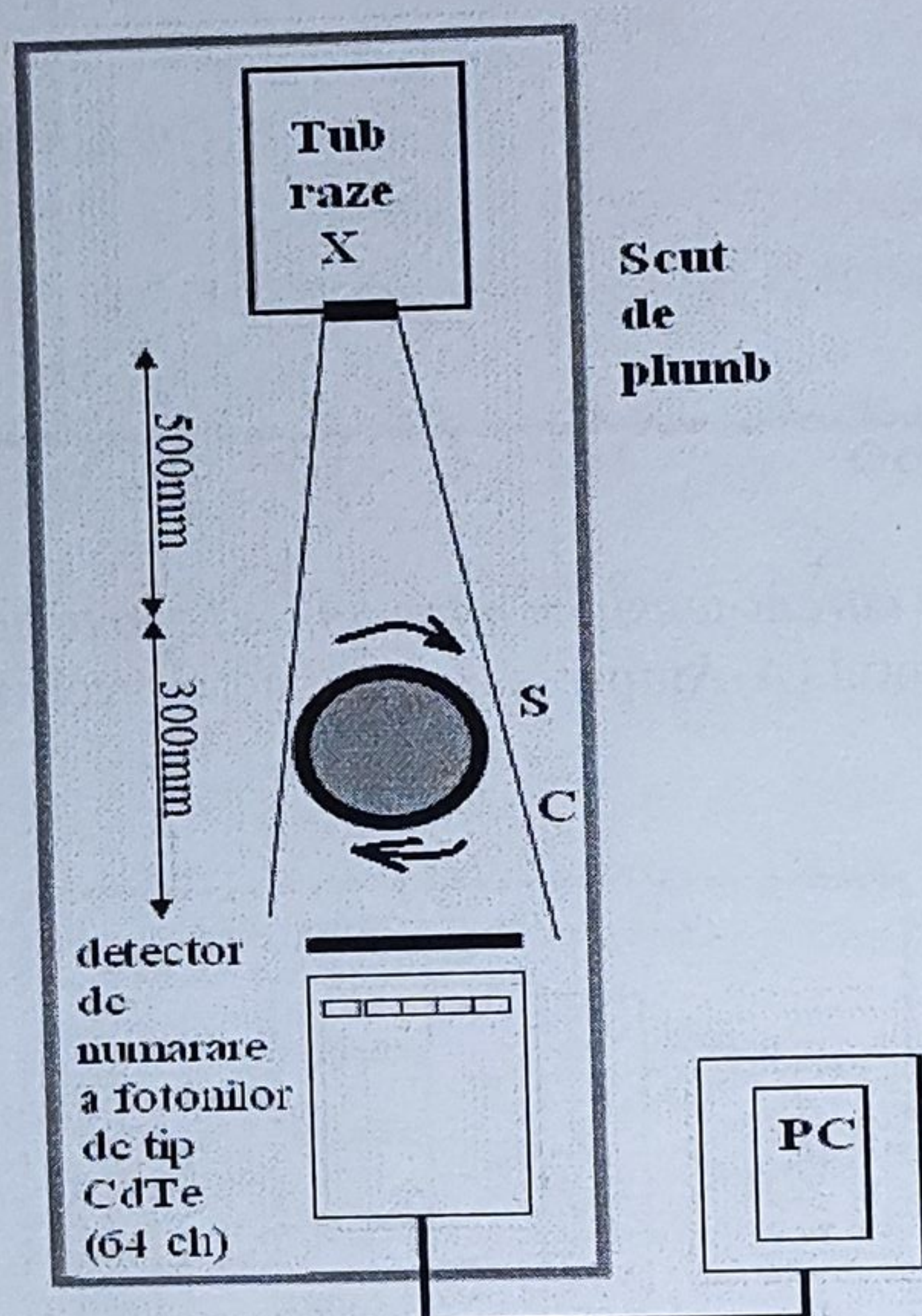


Fig. 13. *Experiment de măsurare a numărului atomic și a densității electronice*

Figura 13 prezintă un sistem experimental pentru determinarea numărului atomic și măsurarea densității de electroni. Acest sistem a fost plasat într-un ecran de plumb și echipat cu un tub de raze X convențional, un manipulator XYθ un colimator, detectorul de CdTe pentru numărarea fotonilor, și un calculator. Colimatorul a fost făcut dintr-o placă de molibden cu o grosime de 2.0 mm cu 64 de găuri de diametru de 0,6 mm aliniate una lângă alta, la distanțe egale de 1,0 mm. Colimatorul are două funcții importante. Prima este aceea de colimare a fasciculului de raze X pentru a restrânge măsurarea transmisiei la o „felie” cu o grosime de câțiva milimetri. A doua funcție importantă constă în reducerea unghiului de împrăștiere a radiației la câteva grade, scăzând astfel efectul împrăștierii.

¹⁴ Yasuhiro Tomita, Yuji Shirayanagi, Masaki Misawa, Tooru Aoki, and Yoshinori Hatanaka, *X-ray Color Scanner with Multiple Energy Discrimination Capability*, Proc. SPIE 5922 (2005) 59220A.

Deoarece detectorul operează în modul de numărare a fotonilor, se obține diferența energiei în câmpul de raze X detectat. Tensiunea și curentul tubului de raze X au fost stabilite la 140 keV, respectiv la $10 \mu A$. Deoarece rezoluția energiei detectorului de CdTe este de aproximativ 10 keV, la 122 keV, lărgimea fiecărei benzi de energie a fost fixată la 20 keV. Pentru a reduce zgomotul, cea mai mică valoare a tensiunii de prag a fost aleasă cea corespunzătoare fotonilor radiației X cu energia de 50eV. Pentru alte trei tensiuni de prag, acestea au fost setate astfel încât să corespundă valorilor energiei de 70, 120 și 140 keV. Astfel, datele măsurate în două intervale energetice, 50-70 și 120-140keV pot fi folosite pentru reconstrucție. Pentru o evaluare cantitativă a numărului atomic și a densității de electroni au fost efectuate măsurători pe diverse probe. Deoarece dinamica numărului atomic și densitatea de electroni necesare pentru un scanner CT variază de la aer la obiectele din metal, am luat ca probe materiale elementale de puritate mare (C, Mg, Al și Ti). Datele de proiecție pot fi reconstruite pentru a da o distribuție a coeficientului de atenuare liniară utilizând tehnica de reconstrucție CT. Am folosit ca filtru algoritmul de proiecție invers pentru procesul de reconstrucție. Deoarece acest proces este doar o operație matematică, valoarea reconstruită μ_{CT} nu reprezintă coeficientul de atenuare liniară absolut.

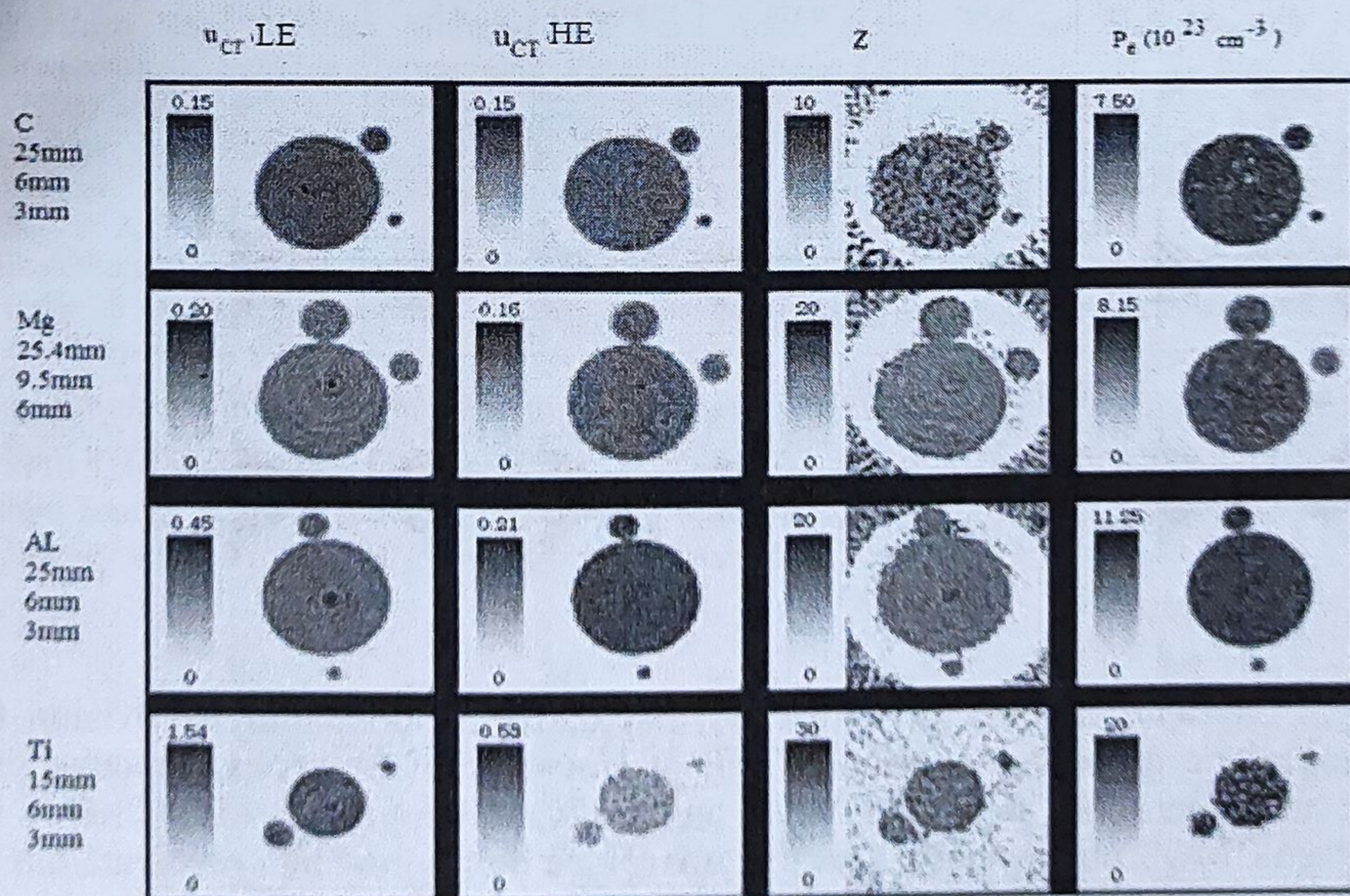


Fig. 14. Imaginile distribuției bazate pe μ_{CT} a două benzi energetice diferite (atât la energie scăzută cât și la energie înaltă), numărul atomic dedus și densitatea de electroni pentru probe de C, Mg, Al și Ti cu diametre diferite

Teoretic, coeficientul de atenuare liniară ar trebui să aibă o relație liniară cu μ_{CT} corespunzător. Figura 14 prezintă imagini ale distribuției bazate pe μ_{CT} pentru

două benzi de energie diferite, numărul atomic dedus și densitatea de electroni a probelor de C, Mg, Al și Ti. Cu scopul de a testa efectele cauzate de adâncimea de penetrare cu raze X, fiecare măsurătoare a fost efectuată pentru trei probe compuse din același element, dar cu diametre diferite.

Tabelul 2 arată datele experimentale citite din figura 14 pentru fiecare material elemental cu diametre diferite. Sunt prezentate și rezultatele teoretice pentru densitatea de electroni pentru fiecare element. Pentru materialele cu număr atomic Z mic, Tabelul 2 demonstrează că separarea materialului poate fi obținută plecând de la imaginile distribuției numărului atomic și a densității de electroni.

Diferența dintre numerele atomice măsurate și valorile teoretice are valori întotdeauna mai mici decât 1.

	ϕ mm	μ_{CT} Low energy	μ_{CT} High energy	Z^{exp}	$ (Z^{exp} - Z^{th})/Z^{th} $ %	ρ_e^{exp} 10^{23} cm^{-3}	$ (\rho_e^{exp} - \rho_e^{th})/\rho_e^{th} $ (%)
C $Z^{th} = 6, \rho_e^{th} = 6.8 \times 10^{23} \text{ cm}^{-3}$	25	0.120	0.106	6.53 ± 0.03	8.8	5.89 ± 0.01	13.4
	6	0.127	0.116 ± 0.001	6.42 ± 0.05	7	6.21 ± 0.05	8.7
	3	0.127 ± 0.001	0.115 ± 0.001	6.08 ± 0.04	0.5	6.25 ± 0.03	8.1
Mg $Z^{th} = 12, \rho_e^{th} = 5.2 \times 10^{23} \text{ cm}^{-3}$	25.4	0.178	0.108	11.27 ± 0.03	6.1	5.74 ± 0.02	10.4
	9.5	0.176	0.106 ± 0.001	11.57 ± 0.08	3.6	5.60 ± 0.01	7.7
	6	0.177	0.104 ± 0.001	11.95 ± 0.06	0.4	5.42 ± 0.02	4.2
Al $Z^{th} = 13, \rho_e^{th} = 7.8 \times 10^{23} \text{ cm}^{-3}$	25	0.305	0.170	12.39 ± 0.01	4.7	8.83 ± 0.01	13.2
	6	0.299	0.169 ± 0.001	12.56 ± 0.02	3.4	8.81 ± 0.01	12.9
	3	0.288 ± 0.001	0.155 ± 0.003	12.71 ± 0.18	2.2	8.10 ± 0.04	3.8
Ti $Z^{th} = 22, \rho_e^{th} = 12.5 \times 10^{23} \text{ cm}^{-3}$	15	1.122 ± 0.002	0.387 ± 0.002	18.3 ± 0.2	16.8	17.5 ± 0.3	40
	6	1.144 ± 0.006	0.388 ± 0.005	18.5 ± 0.2	15.9	16.7 ± 0.2	34
	3	1.216 ± 0.020	0.359 ± 0.014	20.6 ± 0.2	6.4	16.0 ± 0.3	28

Tabelul 2. Datele experimentale atașate figurii 14, pentru fiecare material elemental - probe cu diverse diametre. Sunt notate și rezultatele teoretice ale densității de electroni pentru fiecare element.

Concluzii

Capitolul descrie un număr de dispozitive în domeniul nanovision (NV) dezvoltate de grupul de excelență COE al Universității Shizuoka, Japonia, printre care: un senzor de imagine CMOS cu rezolvare temporală, (TR-CMOS, time-resolved CMOS image sensor), pentru imagistica de fluorescență rezolvată temporal (TD-FLI, time-domain fluorescence lifetime imaging), un microscop optic asistat prin excitație cu un fascicul de electroni și un senzor de imagine în domeniul razelor X cu CdTe, capabil de a identifica numărul atomic și densitatea de electroni.

A fost implementat prototipul senzorului de imagine CMOS utilizând pixeli DOM. S-a reușit măsurarea timpului de dispariție a fluorescenței cu o rezoluție de ordinul sub-nanosecondelor. Senzorul prototip este util pentru realizarea unei camere video compacte cu costuri reduse, pentru imagini microscopice prin măsurarea timpului

de scădere a fluorescenței FLIM utilizabilă în măsurători biologice. Microscopia optică asistată prin excitație cu fascicul de electroni, a fost propusă ca un nou tip de microscopie optică în câmp apropiat. Imaginea luminescenței celulelor vii în soluția de cultură a fost obținută fără limita difracției luminii. Microscopul EXA este un instrument util pentru observarea de celule biologice vii în condiții fiziologice.

Cu scopul de a aplica tehnologia utilizării a două tipuri de energii (dual-energy) pentru identificarea unui material, un nou sistem de scanare CT a fost propus prin utilizarea unui tub de raze X convențional și a unui detector de tip CdTe pentru numărarea fotonilor. Acest sistem poate oferi informații referitoare la datele de proiecție pentru două benzi de energie distincte pentru materialele scanate. Numărul atomic corespunzător și densitatea de electroni pot fi deduse prin metoda de CT cu raze X și energie dublă, adoptată. Prin compararea rezultatelor obținute cu cele teoretice, s-a demonstrat fezabilitatea utilizării acestui sistem pentru identificarea unor materiale cu număr atomic Z mic.

Bibliografie

- Hidenori Mimura, Kazuhiko Hara, Shoji Kawahito, Toru Aoki, and Norihisa Hiromoto, *Nanovision Science-Evolution of Imaging Technology*, Corona Pub. Co. Ltd., (2009) in Japanese.
- Zhuo Li, Shoji Kawahito, Keita Yasutomi, Keiichiro Kagawa, Juichiro Ukon, and Hirohiko Niioka, *A Time-Resolved CMOS Image Sensor With Draining-Only Modulation Pixels for Fluorescence Lifetime Imaging*, IEEE Trans. Electron Devices, 59 (2012) 2715.
- Wataru Inami, Kentaro Nakajima, Atsuo Miyakawa, and Yoshimasa Kawata, *Electron Beam Excitation Assisted Optical Microscope with Ultra-High Resolution*, Optics Express 18 (2010) 12897.
- Wenjuan Zou, Takuya Nakashima, Yoshiaki Onishi, Akifumi Koike, Bunji Shinomiya, Hisashi Morii, Yoichiro Neo, Hidenori Mimura, and Toru Aoki, *Atomic Number and Electron Density Measurement Using a Conventional X-ray Tube and a CdTe Detector*, Jpn. J. Appl. Phys. 47 (2008) 7317.
- Hyung-June Yoon, Shinya Itoh, and Shoji Kawahito, *A CMOS Image Sensor with in-Pixel Two-Stage Charge Transfer for Fluorescence Lifetime Imaging*, IEEE Transactions on Electron Devices, Vol. 56, Nr. 2, (Febr. 2009) 214.
- Yumiko Ohno, Masami Torikoshi, Takanori Tsunoo, and Kazuyuki Hyodo, *Dual-energy X-ray CT with CdTe Array and its extension*, Nuclear Instruments and Methods in Physics Research Section A548, pp 72-77, (2004).
- Toru Aoki, Y. Ishida, D. Sakashita, V. Gnatyuk, A. Nakamura, Y. Tomita, Y. Hatanaka, and J. Tenmmyo, *Development of Energy-Discriminate CdTe Imaging Detector for Hard X-ray*, Proc. SPIE 5540, 196, 2004.
- Yasuhiro Tomita, Yuji Shirayanagi, Masaki Misawa, Tooru Aoki, and Yoshinori Hatanaka, *X-ray Color Scanner with Multiple Energy Discrimination Capability*, Proc. SPIE 5922 (2005) 59220A

Stabilirea corespondențelor între pixeli în stereoviziunea binoculară

ALAIN CROUZIL¹, GUILLAUME GALES, SYLVIE CHAMBON

ABSTRACT. Stereo matching is one of the main topics in computer vision. This is an essential step for 3D reconstruction by binocular stereo vision. It consists in finding in two images of a same scene, taken from different viewpoints, the pairs of pixels which are the projections of a same scene point. Since the last twenty years, many local and global methods have been proposed to solve this problem. More recently, region-based methods showed interesting result in small-baseline binocular stereo (where images are taken nearby). This chapter describes the main approaches of stereo matching. A first part presents the local methods that search for correspondents of each pixel independently assuming that the neighbors of two corresponding pixels are similar. The next part describes a specific local method based on seed propagation, i.e. which is based on an initial set of reliable correspondences, reducing the size of the search area for matching their neighbors. Then, another part deals with the global methods that search the correspondent of all the pixels by minimizing a global error over the whole image. Finally, the last part describes the methods that rely on a segmentation of images giving regions for which the parameters of a surface model are estimated.

KEYWORDS: *stereo matching, binocular stereo vision, small-baseline, segmentation, disparity map.*

1. Introducere

În încercarea de a reproduce funcționalitățile percepției vizuale umane, domeniul viziunii computerizate a condus la noi tematici de cercetare. S-au realizat sisteme automate care, pornind de la una sau mai multe imagini, percep sau măsoară relieful sau mișcarea, recunosc obiecte, scene, situații. Aceste lucrări contribuie la interpretarea scenei pentru a regăsi, a localiza, identifica și caracteriza ceea ce se află sau ceea ce se întâmplă în cadrul scenei fotografiate sau filmate.

¹ Profesorul Alain Crouzil – IRT, Institutul de Cercetări în Informatică din Toulouse, cadru didactic la Universitatea Paul Sabatier, Toulouse, Franța, co-autor a acestui capitol, colaborare dezvoltată în stagiul post-doctoral efectuat de Mihaela Costin în cadrul proiectului POSDRU 56815, „Societatea bazată pe cunoaștere, cercetări, dezbateri, perspective”, la IRT, Toulouse, Franța. Capitolul reprezintă o primă publicare și traducere a părților comune, nepublicate, din tezele și rapoartele de cercetare conduse de Dl. Prof. Alain Crouzil.

1.1 Două camere de luat vederi pentru percepția reliefului

Prin analogie cu sistemul vizual uman, stereoviziunea binoculară utilizează o pereche de imagini ale aceleiași scene captate din poziții diferite. Cele două camere de luat vederi se bazează pe principiul vederii binoculare umane care captează razele luminoase ce provin de la scenă, iar poziția diferită permite perceperea adâncimii. Algoritmii de stereoviziune binoculară au ca obiectiv să calculeze poziția în cadrul scenei, și în particular, profunzimea elementelor conținute în cele două imagini. Adeseori, aceste elemente sunt puncte, dar unele lucrări, în particular cele din robotică, iau în considerare în aceeași măsură și segmentele de dreaptă.

În general, sunt necesare patru etape mari pentru a extrage informația de profunzime prin intermediul stereoviziunii binoculare:

- **Achiziția imaginilor** – cele două imagini ale aceleiași scene sunt captate în același timp de două camere de luat vederi diferite sau, în cazul unei scene fixe, poate fi utilizată o singură cameră care achiziționează cele 2 imagini din poziție diferită.
- **Calibrarea sistemului de captare a imaginilor** – această etapă constă în determinarea parametrilor interni și externi ai modelului geometric al sistemului de captare a imaginilor (conform Fig. 1). În marea majoritate a cazurilor, se adoptă modelul camerei pinhole pentru care formarea geometrică a imaginilor constă într-o proiecție în perspectivă. Numeroase lucrări au avut ca subiect calibrarea, ca de exemplu [Bocquillon, 2008]².
- **Stabilirea corespondențelor între pixeli** – se referă la regăsirea în cele două imagini a perechilor de pixeli ce corespund proiecției aceluiași element din scenă (conform Fig. 1).
- **Reconstrucția 3D** – constă în calcularea pentru fiecare pixel a poziției în spațiu a punctului care este proiectat pe acel pixel. Această etapă, denumită triangularizare, necesită cunoașterea corespondențelor rezultate de la etapa precedentă.

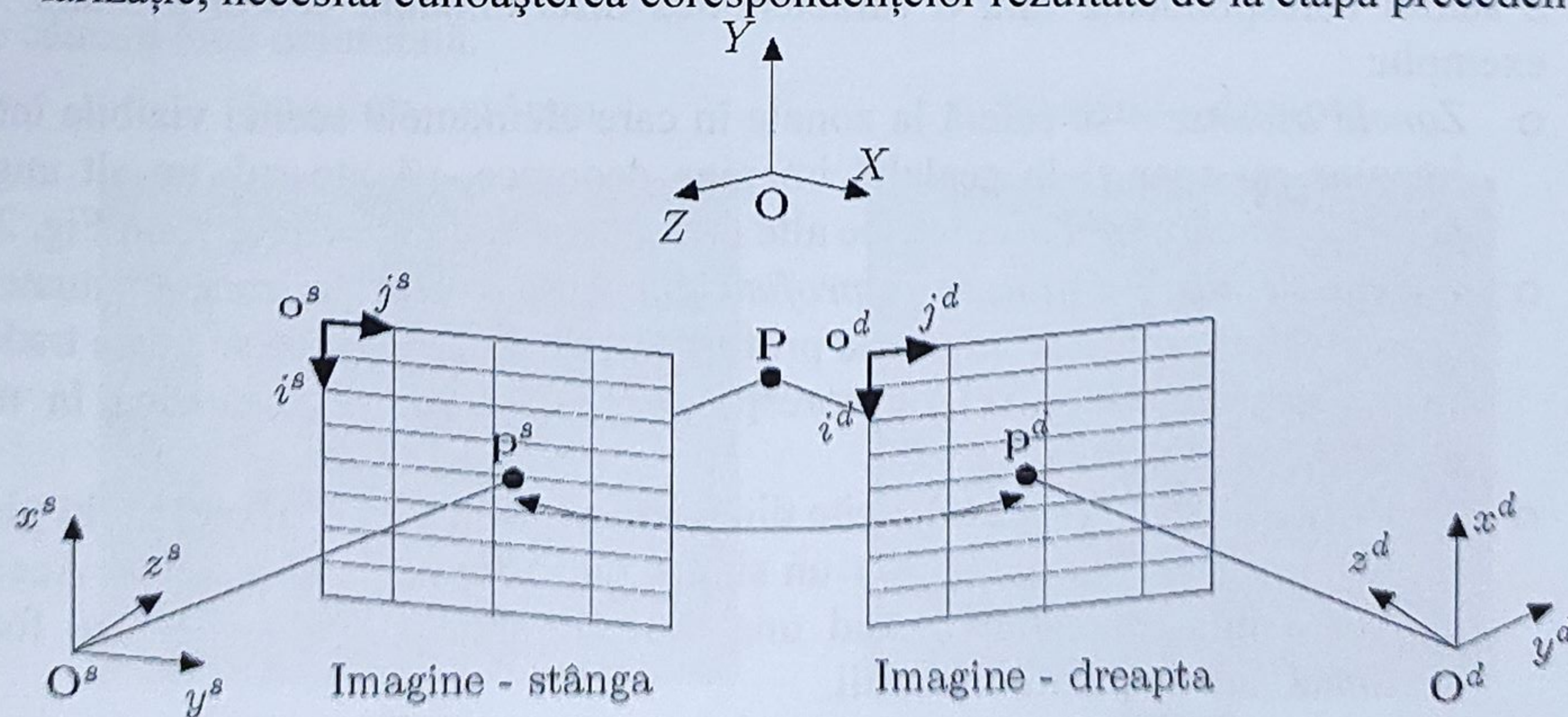


Fig. 1. Modelul geometric al sistemului de captare stereoscopică binoculară

Acest capitol vizează etapa de punere în corespondență a pixelilor în cazul stereoviziunii pentru care distanța între centrele celor două camere este mică (small-

² Bocquillon, Benoît, „Contributions à l'autocalibrage des caméras: modélisations et solutions garanties par l'analyse d'intervalle”, teză de doctorat, Universitatea din Toulouse, Oct. 2008.

baseline). În figura 1 este prezentat modelul geometric al sistemului de captura stereoscopică binoculară, în care pixelii notați cu p^s și p^d corespund unul cu celălalt pentru că sunt proiecțiile aceluiasi punct P din scenă pe imaginile din stânga și din dreapta. Reperele (O^s, x^s, y^s, z^s) și (O^d, x^d, y^d, z^d) sunt reperele camerei de luat vederi din stânga și, respectiv, din dreapta. Reperul (O, X, Y, Z) este reperul scenei. Reperele imaginilor sunt notate cu (o^s, i^s, j^s) și (o^d, i^d, j^d) .

1.2 Stabilirea corespondențelor între pixeli

Obiectivul punerii în corespondență a pixelilor în cazul stereoviziunii binoculare constă în regăsirea în cele două imagini a perechilor de pixeli care corespund proiecției unei singure entități din scenă. Este vorba despre o problemă care a făcut obiectul a numeroase lucrări de cercetare deoarece ea prezintă mai multe dificultăți pe care le vom detalia în cele ce urmează, spre deosebire de etapa de calibrare, pentru care există posibilitatea de a pune în practică metode precise și eficiente. În ciuda eforturilor cercetătorilor, abordările propuse încearcă doar să depășească anumite dificultăți, dar nu ajung să rezolve simultan toate problemele. De cele mai multe ori este necesar să se recurgă la un compromis.

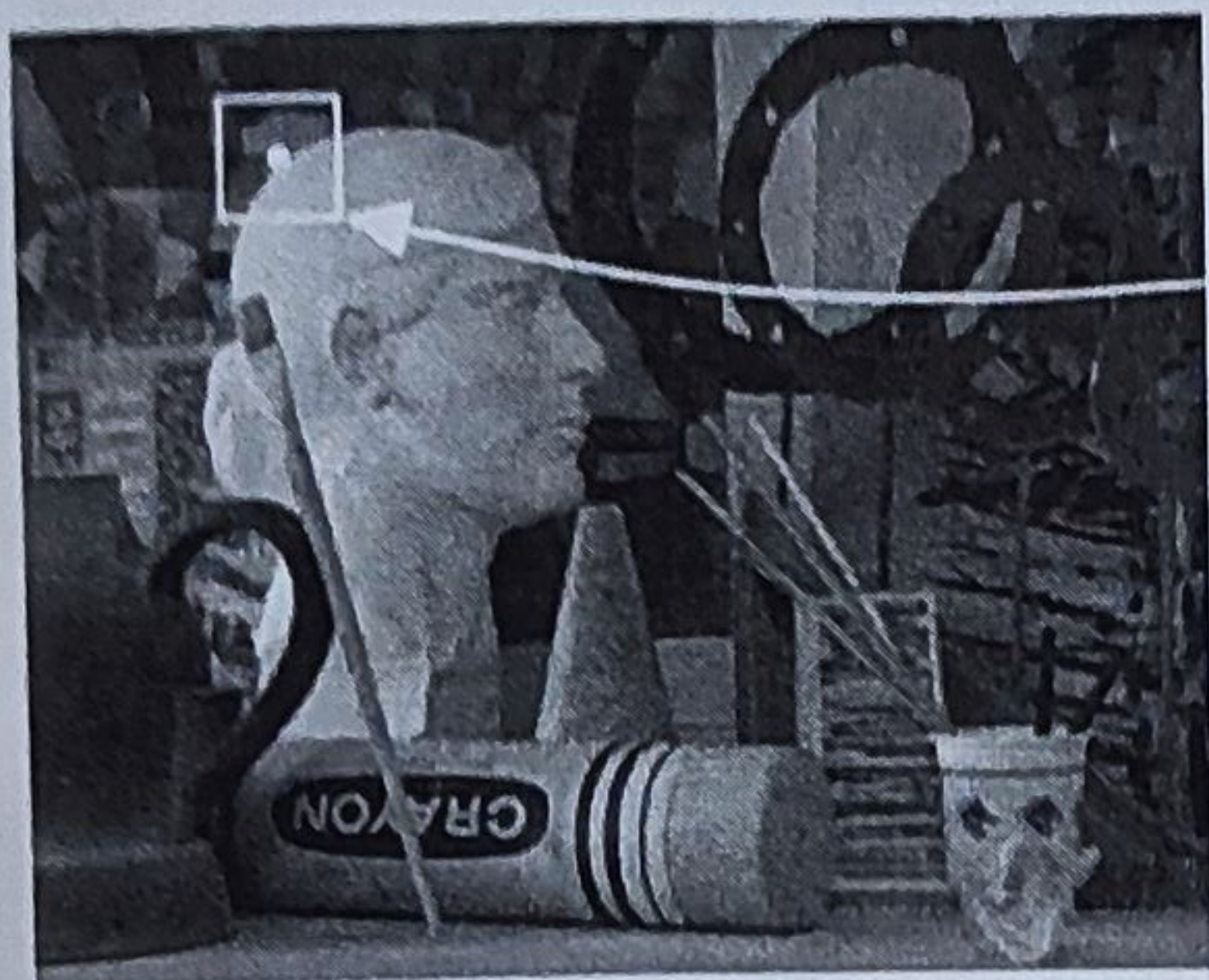
1.3 Câteva dificultăți

Metodele de punere în corespondență a pixelilor se bazează în general pe următoarea ipoteză: vecinătățile a doi pixeli corespondenți se aseamănă. Principalele dificultăți întâmpinate în timpul punerii în corespondență au în esență două rațiuni:

- **Informații care lipsesc** – informațiile care ar permite realizarea unei puneri în corespondență de manieră sigură nu sunt prezente în imagine. Este deci dificil de a stabili corespondențe fără o caracteristică discriminantă complementară. De exemplu:
 - *Zonele ascunse* – se referă la zonele în care elementele scenei vizibile într-o imagine nu apar și în cealaltă imagine deoarece, văzute sub un alt unghi, aceste elemente sunt ascunse de alte elemente din imagine (conform Fig. 2).
 - *Zonele de discontinuitate a profunzimii* – se referă la zonele situate la marginea unor obiecte situate la profunzimi diferite. Ceea ce se poate traduce prin vecinătăți ale pixelilor corespondenți care nu se aseamănă în mod necesar (conform Fig. 3).
 - *Zonele de tăiere* – puncte diferite din scenă se proiectează în mai mulți pixeli dintr-o imagine, dar numai într-un singur pixel din cealaltă imagine. Aceasta se poate întâmpla atunci când una din suprafețele din scenă este foarte înclinată față de planul imaginii.
- **Informații ambigue** – informația este ambiguă dacă este prezentă în imagine dar nu permite ca să se distingă pixelii corespondenți.
 - *Zone omogene* – pixelii unei zone omogene se aseamănă între ei, în cazul unei imagini zgomotoase. În consecință, este foarte dificil de făcut distincție între aceștia (conform Fig. 4).

- *Zone de textură regulată* – există o ambiguitate dacă un model se repetă identic (conform Fig. 5).

Imagine - stânga



Imagine - dreapta

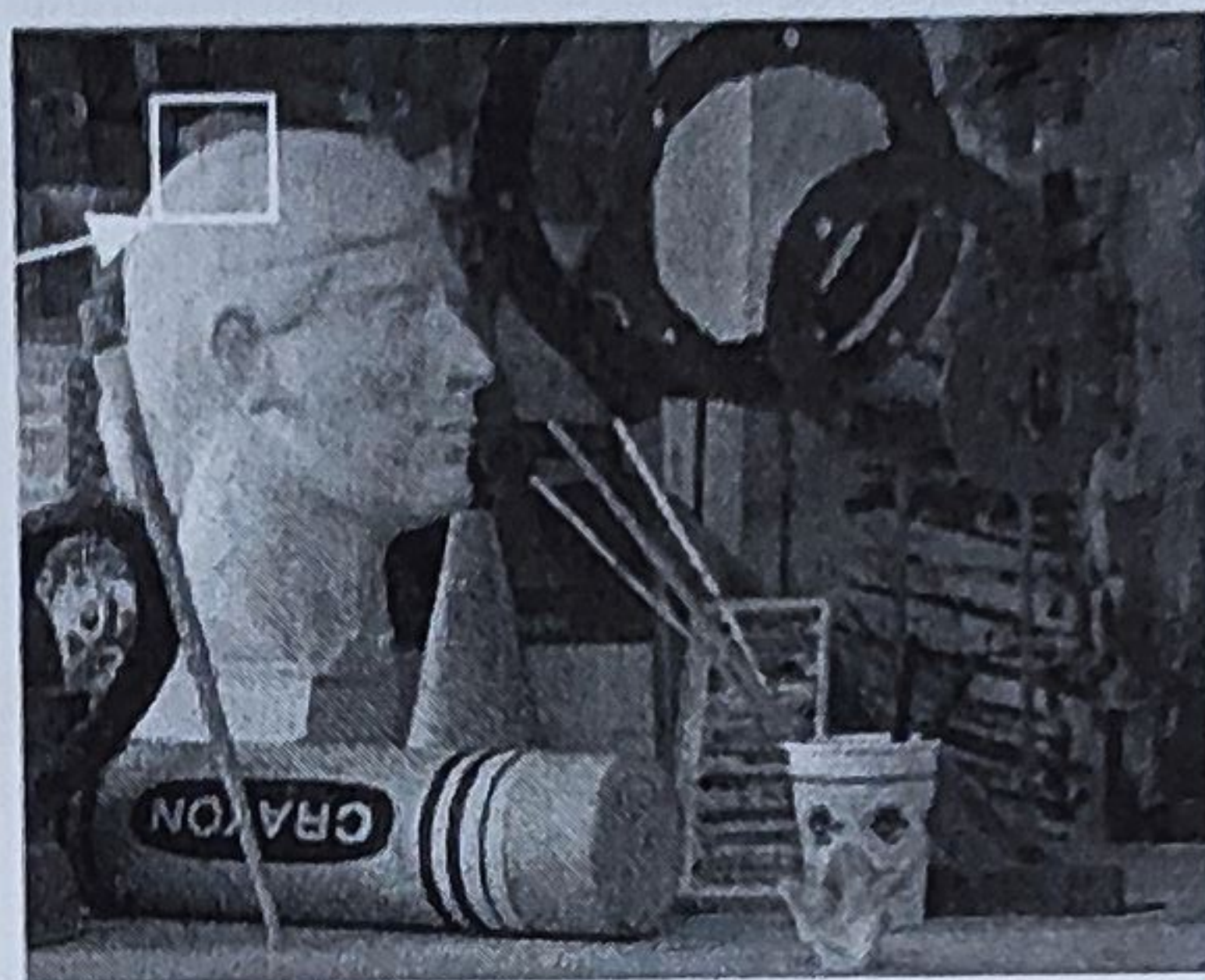
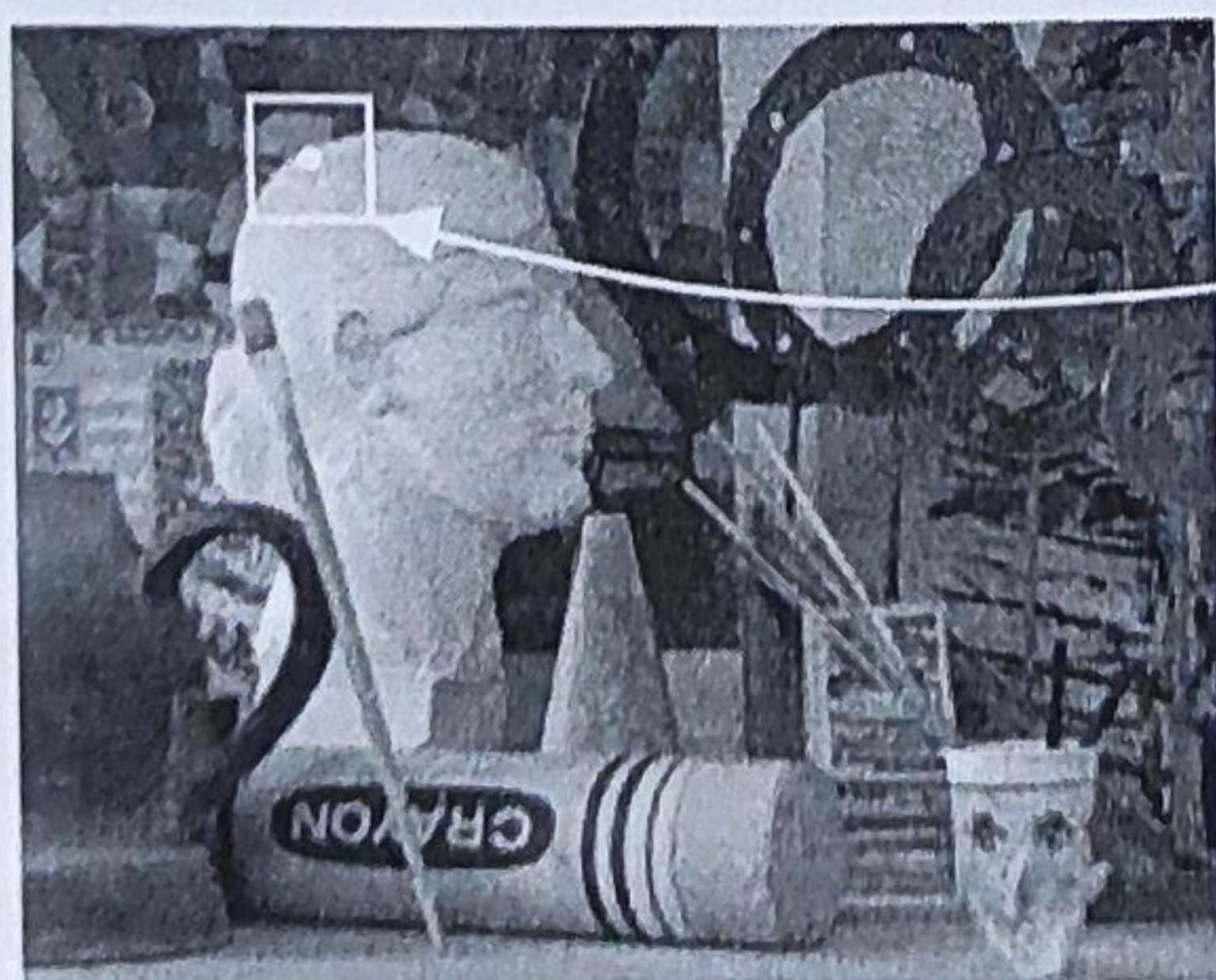


Fig. 2. *Zone ascunse* – pixelul marcat cu un punct alb are corespondentul ascuns în imaginea din dreapta

1.4 Căi de abordare

Metodele de punere în corespondență a pixelilor se pot împărți în două grupe. Prima grupă este cea a metodelor globale care iau în considerare totalitatea imaginii și încearcă să minimizeze o funcție de cost care exprimă erorile de punere în corespondență. Cea de-a doua grupă este cea a metodelor locale în care perechile de pixeli corespondenți sunt găsite prin măsurarea similarității vecinătăților lor. Corespondentul fiecărui pixel este căutat dintr-o serie de candidați situați într-o zonă de căutare bine delimitată.

Imagine - stânga



Imagine - dreapta

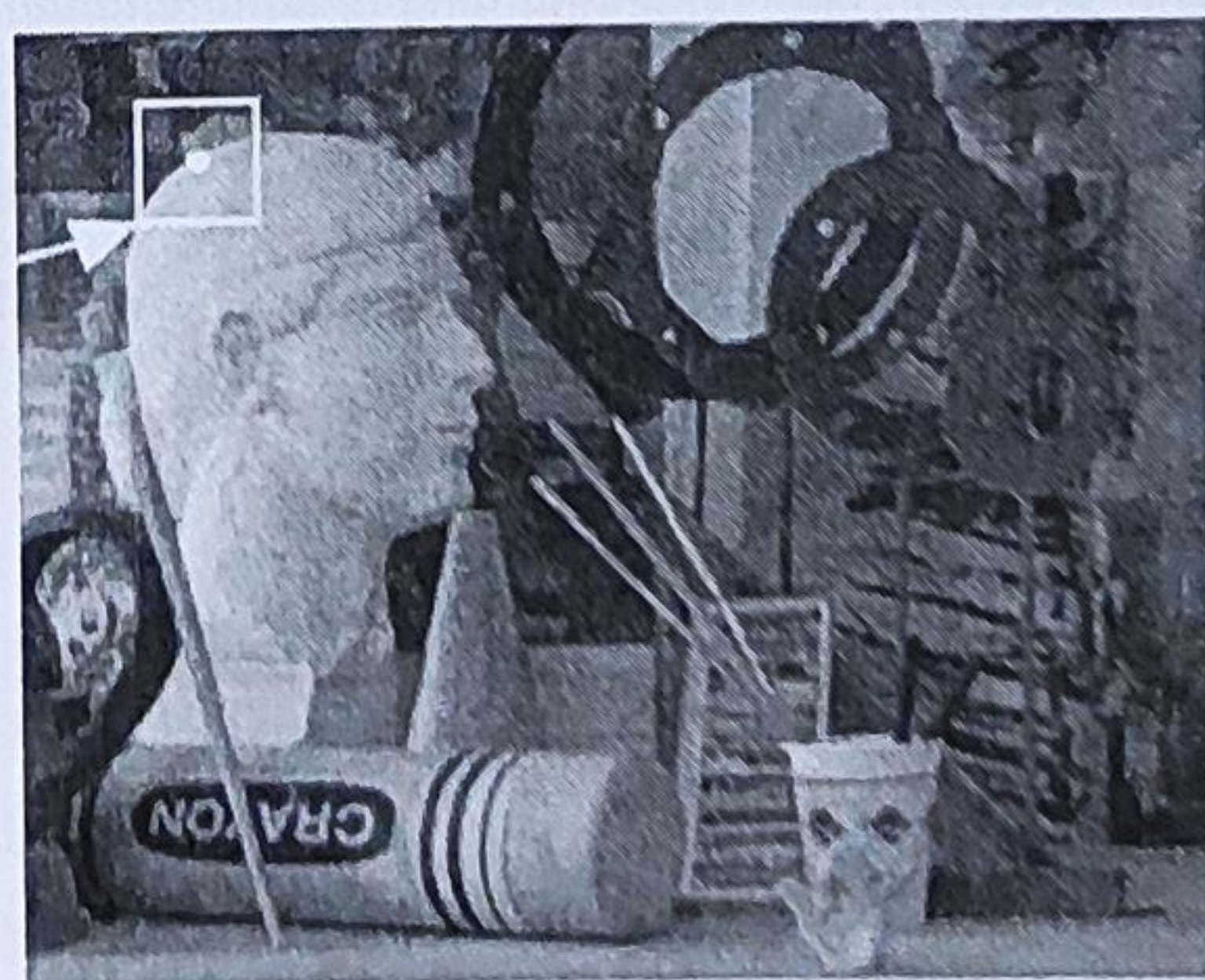


Fig. 3. *Zone de discontinuitate a profunzimii* – cei doi pixeli marcați cu un punct alb corespund unul cu celălalt dar vecinătățile lor diferă pentru că sunt apropiați de o discontinuitate de profunzime

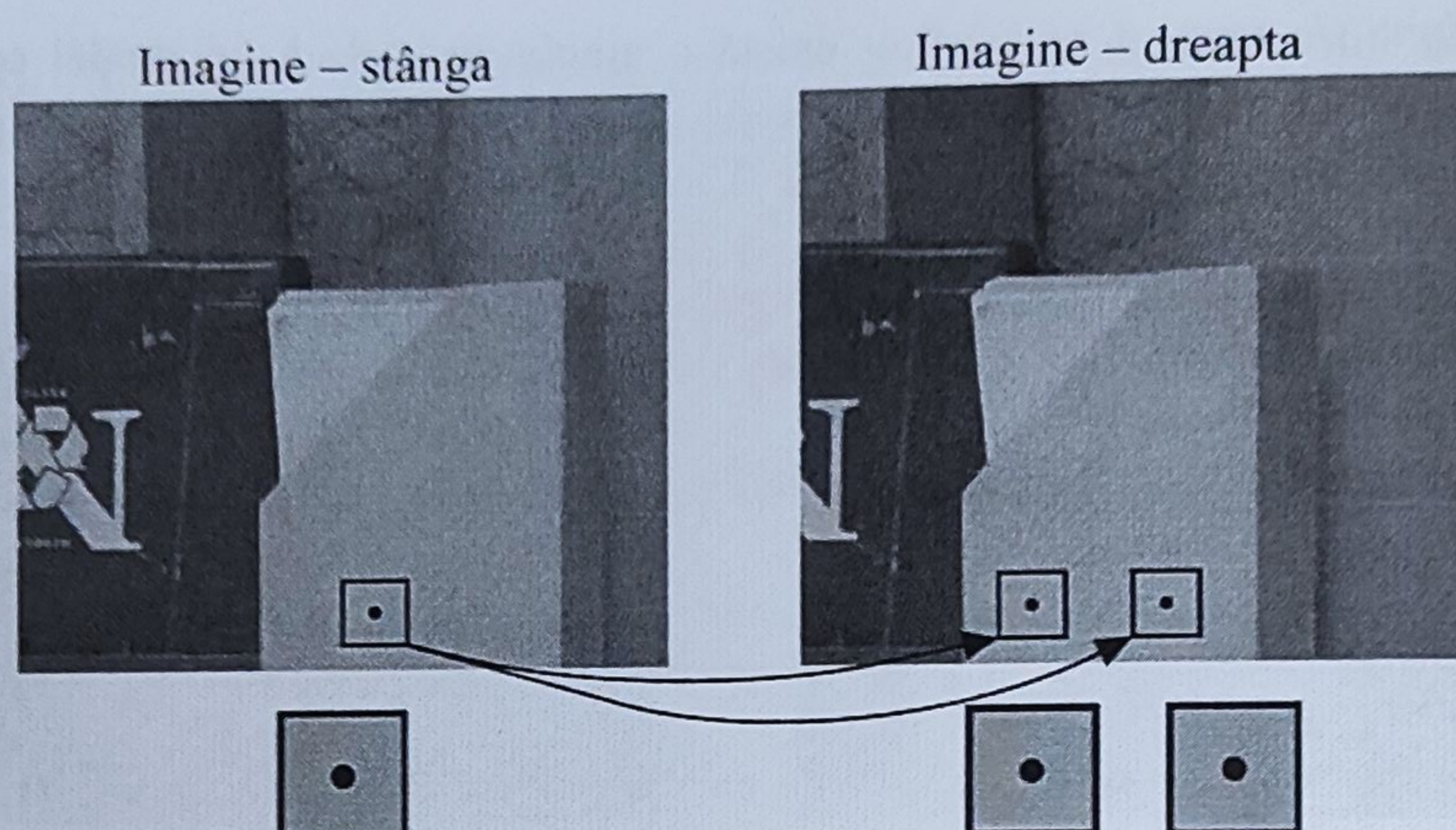


Fig. 4. *Zone omogene* – în zonele omogene nu avem prea multă informație pentru a putea caracteriza pixelii; de aceea este dificil de găsit pixelul corespondent.

Metodele locale pot fi implementate cu oarecare ușurință și sunt din ce în ce mai utilizate în aplicații diferite. Din nefericire, ele sunt sensibile la dificultățile prezentate în paragraful precedent. De aceea, există o subgrupă particulară de metode bazate pe creșterea regiunilor (*propagation de germes*). Ele permit a priori reducerea anumitor ambiguități cu care se confruntă metodele clasice, precum și reducerea timpului de calcul. De altfel, clasamentul oferit de protocolul de evaluare a metodelor de punere în corespondență densă a pixelilor pentru stereoviziunea binoculară a lui Middlebury³ [Scharstein, 2002]⁴ pune în evidență rezultatele foarte bune obținute de metodele bazate pe regiuni. Este vorba despre o familie de metode hibride, deoarece în general ele constau într-o etapă de punere în corespondență locală pentru a inițializa algoritmul, urmată de o etapă de optimizare globală pentru rafinarea rezultatelor. Metodele bazate pe regiuni pornesc de la ipoteza că pixelii din aceeași regiune care rezultă în urma unei segmentări de culoare omogenă sunt proiecții ale unei aceleiași suprafețe. Această ipoteză este în general adevărată aproape peste tot, iar contururile obiectelor din cadrul scenei, acolo unde se situează discontinuitățile de profunzime, coincid cu marginile regiunilor.

1.5 Organizarea capitolului

Abundența de lucrări științifice care fac obiectul punerii în corespondență a pixelilor nu ne permite să fim exhaustivi. Obiectivul acestui capitol este să ofere o privire de ansamblu asupra pistelor care au fost abordate și de a ghida cititorul care dorește să știe mai multe.

Secțiunea 2 este consacrată metodelor locale care caută corespondenții fiecărui pixel într-o manieră independentă presupunând că vecinătățile a doi pixeli corespondenți se aseamănă. Secțiunea 3 prezintă o abordare locală particulară bazată

³ <http://vision.middlebury.edu/stereo/>

⁴ Scharstein, Daniel, Szeliski, Richard, „A taxonomy and evaluation of dense twoframe stereo correspondence algorithms”, *International Journal of Computer Vision*, 47(1):7–42, 2002.

pe propagarea germenilor, care pornește de la un set de pixeli corespondenți fiabili, germenii, permițând limitarea mărimii zonei de căutare a pixelilor corespondenți printre vecinii acestora. Secțiunea 4 tratează metodele globale care caută corespondenții tuturor pixelilor prin minimizarea unei erori globale pe toată imaginea. În sfârșit, secțiunea 5 descrie metodele care utilizează o segmentare a imaginilor în regiuni pentru care sunt estimați parametrii unui model de suprafață.

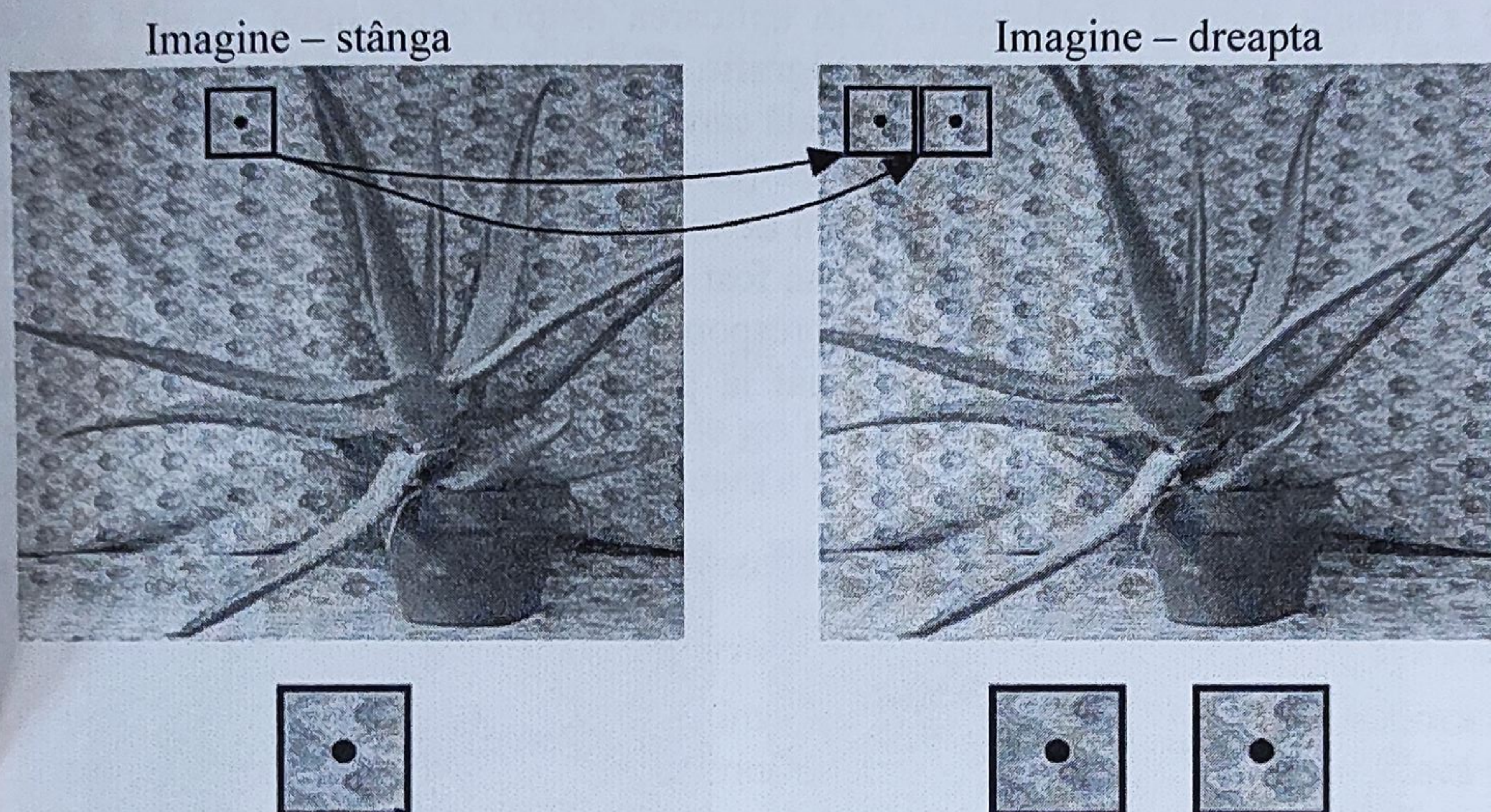


Fig. 5. *Zone cu textură regulată* – pentru un pixel există mai mulți candidați posibili pentru că vecinătățile se aseamănă. Din această cauză este dificil de găsit pixelul care corespunde cel mai bine.

2. Metode locale

Metodele locale clasice caută în cele două imagini pixelii ale căror vecinătăți se aseamănă cel mai mult prin utilizarea măsurilor de corelație. Pentru fiecare pixel dintr-o imagine a cuplului, căutarea pixelului corespondent în cealaltă imagine se face într-o serie de pixeli candidați care se constituie într-o zonă de căutare. Fără cunoștințe a priori, această zonă poate să se întindă la ansamblul pixelilor din cealaltă imagine, dar este posibil ca ea să fie redusă, în particular prin utilizarea modelului geometric al sistemului de captare a imaginilor care introduce condiția de epipolaritate.

2.1 Rectificarea epipolară

Condițiile geometrice permit reducerea ansamblului de pixeli corespondenți posibili, în particular prin utilizarea geometriei epipolare care descrie relațiile dintre pixelii corespondenți. Această relație este descrisă de parametrii incluși în matricea

fundamentală [Hartley, 2003]⁵. Condiția de epipolaritate specifică faptul că pixelul corespondent al unui punct dintr-o imagine se găsește pe o dreaptă din cealaltă imagine. Această dreaptă este dreapta epipolară asociată și ecuația ei poate fi obținută pornind de la coordonatele punctului și pe baza matricei fundamentale.

Una din configurațiile particulare ale sistemului de captare a imaginilor este aceea în care dreptele epipolare sunt toate paralele și orizontale. Există posibilitatea de a simula această configurație prin aplicarea asupra celor două imagini a unor transformări geometrice denumite omografii. Matricele acestor omografii pot fi obținute pornind de la matricea fundamentală care, ea însăși, poate fi estimată pornind de la corespondența unor puncte de interes [Loop, 1999]⁶. Fig. 6 prezintă efectul unei rectificări epipolare asupra unei perechi de imagini. Numeroase metode de punere în corespondență consideră că imaginile au fost deja supuse acestei rectificări epipolare. Aceasta permite reducerea căutării corespondentului unui pixel situat pe linia i și coloana j a unei imagini stânga numai la pixelii de pe aceeași linie din cealaltă imagine, situați la stânga coloanei j .

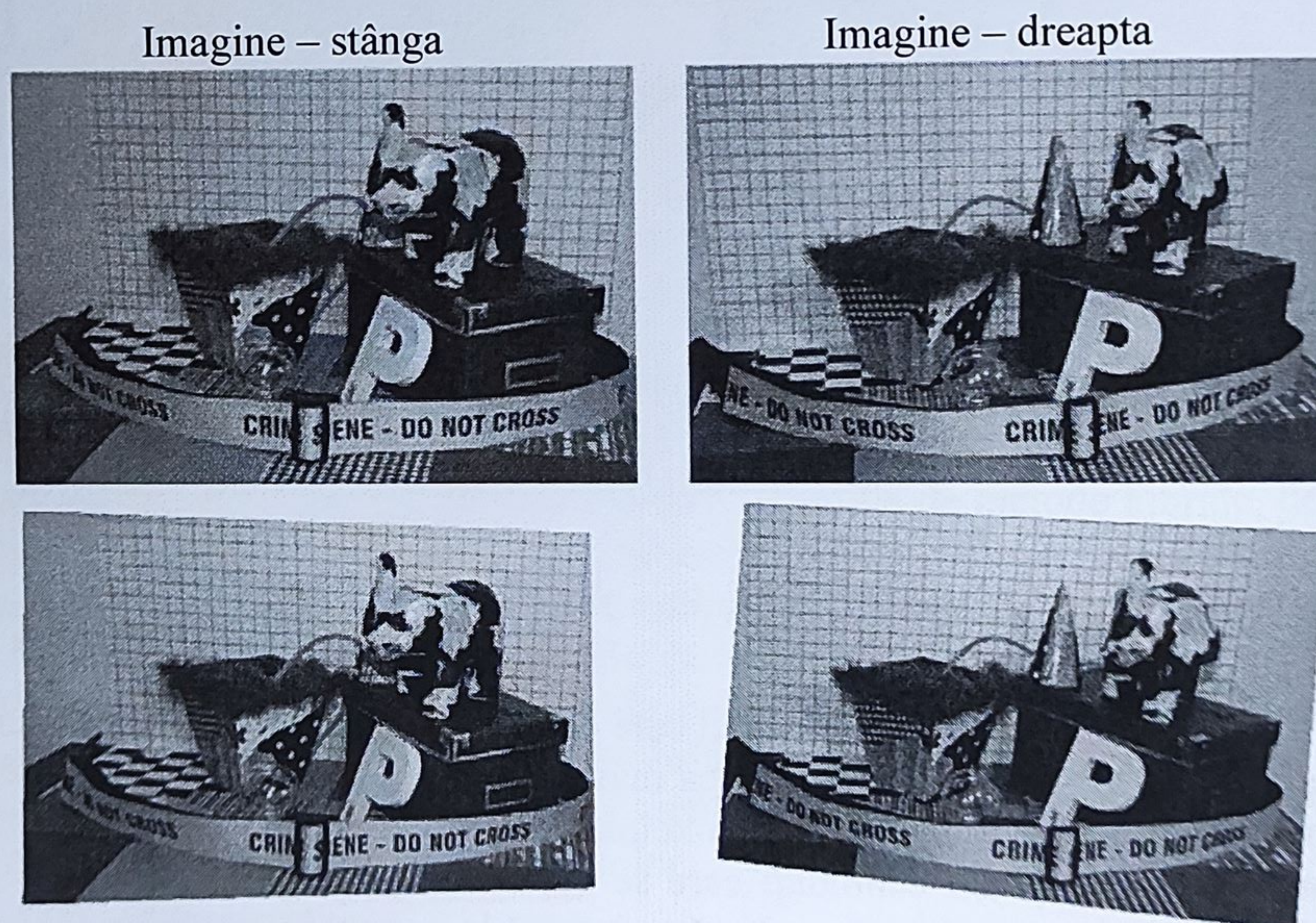


Fig. 6. Rectificarea epipolară – în partea de sus este prezentată o pereche de imagini pentru care s-a efectuat rectificarea epipolară, rezultând imaginile din partea de jos, în care pixelii corespondenți sunt situați pe aceeași linie.

⁵ Hartley, Richard, Zisserman, Andrew, „Multiple view geometry”, Cambridge University Press, second edition, 2003.

⁶ Loop, Charles T., Zhang, Zhengyou, „Computing rectifying homographies for stereo vision”, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pp. 125–131, Fort Collins, U.S.A, June 1999.

2.2 Definiții și notații

Înainte de a prezenta algoritmul elementar al abordării locale, vom preciza mai întâi notațiile care vor fi utilizate în cele ce urmează.

Notăm cu $p_{i,j}^s$, $i \in [0; h]$, $j \in [0; w]$ pixelul situat pe linia i și coloana j a imaginii s de mărime $h \times w$. Notăm cu s imaginea stânga și cu d imaginea dreapta.

- **Funcția imagine I^s** – valoarea unui pixel este dată de „funcția imagine discretizată”:

$$I^s : [0; h] \times [0; w] \rightarrow N^c$$

$$\begin{bmatrix} i & j \end{bmatrix}^T \alpha I^s \left(\begin{bmatrix} i & j \end{bmatrix}^T \right) \quad (1)$$

unde $I^s \left(\begin{bmatrix} i & j \end{bmatrix}^T \right)$ este valoarea dimensiunii c a pixelului $p_{i,j}^s$. Avem, de exemplu, $c = 1$ dacă luăm în considerare nivelele de gri sau $c = 3$ dacă considerăm un spațiu de culoare. Pentru simplificare, vom considera $c = 1$ în continuarea acestui capitol.

- **Vectorul de disparitate $d_{i,j}^{s \rightarrow d}$** – rezultatul punerii în corespondență poate fi reprezentat printr-un *vector de disparitate* $d_{i,j}^{s \rightarrow d}$ care descrie poziția relativă a doi pixeli corespondenți între cele două imagini s și d (conform Fig. 7). Imaginea de pornire, aici s , este imaginea de referință. În continuarea acestui capitol, dacă nu se fac alte precizări, vom considera imaginea stânga ca imagine de referință. Pentru simplificarea notației vom folosi $d_{i,j}^{s \rightarrow d} = d_{i,j}$. Vectorul de disparitate asociat pixelului $p_{i,j}^s$ poate fi descris ca

$d_{i,j} \in ([d_{i_{\min}}; d_{i_{\max}}] \times [d_{j_{\min}}; d_{j_{\max}}]) \cup \{asc\}$. Dacă $d_{i,j} \neq asc$, atunci:

$$\begin{bmatrix} i' \\ j' \end{bmatrix} = \begin{bmatrix} i \\ j \end{bmatrix} + d_{i,j} \quad (2)$$

unde $p_{i',j'}^d$ este pixelul din imaginea dreapta ce corespunde pixelului $p_{i,j}^s$, iar $d_{i_{\min}}, d_{i_{\max}}, d_{j_{\min}}, d_{j_{\max}}$ sunt capetele intervalelor valorilor posibile pentru componentele vectorului de disparitate, în timp ce *asc* este un vector de disparitate simbolic atribuit pixelilor ale căror corespondenți sunt ascunși.

Cum am amintit deja, este posibil ca imaginile să fie rectificate de așa manieră încât corespondentul pixelului din imaginea stângă să se situeze pe aceeași linie în imaginea din dreapta (conform paragrafului 2.1). În acest fel, prima componentă a vectorului de disparitate este întotdeauna nulă. În acest caz, vectorul de disparitate poate fi înlocuit cu o mărime scalară $d_{i,j}$ denumită *disparitate* sau *valoare de disparitate*, exprimată ca:

$$d_{i,j} = \begin{bmatrix} 0 \\ d_{i,j} \end{bmatrix} \quad (3)$$

În continuarea acestui capitol, dacă nu se fac alte precizări, ne vom plasa întotdeauna în acest caz.

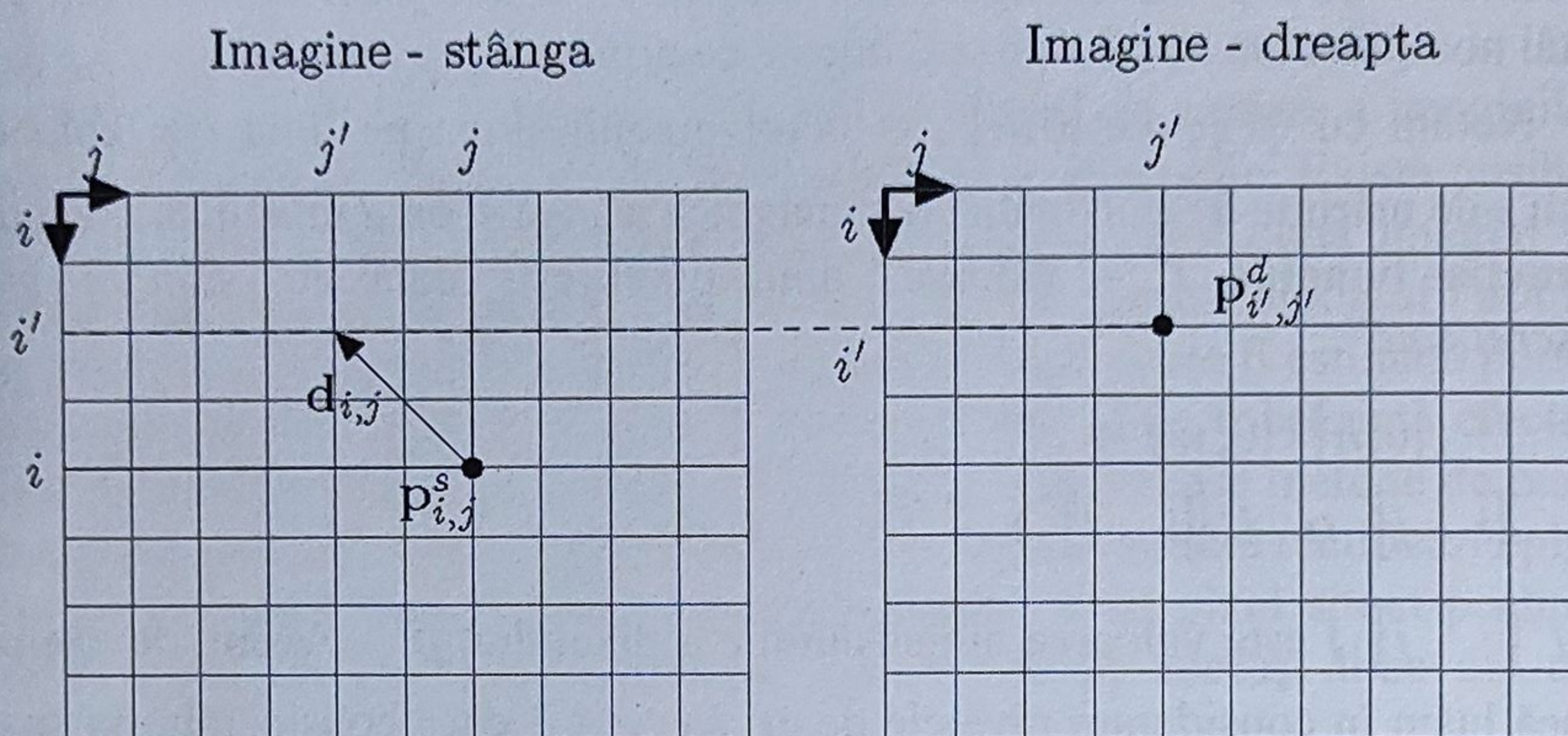


Fig. 7. Vector de disparitate – exprimat ca $d_{i,j} = \begin{bmatrix} i' - i & j' - j \end{bmatrix}^T$

- **Funcția de disparitate d** – câmpul vectorilor de disparitate al unei perechi de imagini poate fi reprezentat printr-o *funcție de disparitate*:

$$d : [0; h] \times [0; w] \rightarrow ([d_{i_{\min}}; d_{i_{\max}}] \times [d_{j_{\min}}; d_{j_{\max}}]) \cup \{asc\}$$

$$\begin{bmatrix} i \\ j \end{bmatrix} \propto d_{i,j} \quad (4)$$

- **Funcția de corespondență m** – la fel putem defini o *funcție de corespondență* care dă coordonatele pixelilor corespondenți:

$$m : [0; h] \times [0; w] \rightarrow ([0; h] \times [0; w]) \cup \phi$$

$$\begin{bmatrix} i \\ j \end{bmatrix} \propto \begin{bmatrix} i' \\ j' \end{bmatrix} \quad (5)$$

- **Hartă de disparitate** – în final, rezultatul punerii în corespondență a pixelilor poate fi vizualizat cu ajutorul unei *hărți de disparitate*. Este vorba despre o imagine pentru care nivelul de gri al fiecărui pixel este proporțional cu disparitatea sa față de imaginea de referință. Cu cât este mai deschis, cu atât elementul care se proiectează pe acel pixel este mai apropiat de sistemul de captare a imaginilor (conform Fig. 8).



Fig. 8. O pereche de imagini și harta de disparitate – în cadrul hărții de disparitate, pixelii negri sunt pixelii ale căror corespondenți sunt ascunși sau necunoscuți

2.3 Algoritmul elementar

Metodele locale măsoară similitudinea între vecinătățile celor doi pixeli. Această vecinătate este denumită *zonă de agregare* sau *fereastră de corelație*. Pentru simplificare, considerăm o vecinătate pătrată ($n \times n = (2r + 1) \times (2r + 1), r \in \mathbb{N}^*$), centrată pe pixelul luat în considerare.

Notăm cu a vectorul de dimensiune n^2 care conține valorile pixelilor vecini pixelului considerat în cadrul imaginii de referință și cu b vectorul similar din cealaltă imagine. Metodele locale presupun că vecinătățile celor doi pixeli corespondenți se aseamănă, adică:

$$a \approx b \quad (6)$$

Această similaritate este evaluată pentru fiecare pixel din imaginea de referință și fiecare pixel din zona acestuia de căutare din cealaltă imagine. În Fig. 9, pixelul din imaginea din stânga, căruia trebuie să-i găsim corespondentul, este marcat cu un punct negru. Corespondentul său este căutat printre pixelii situați în zona de căutare. În acest exemplu, zona de căutare este unidimensională. Ea este definită în funcție de poziția pixelului din stânga pentru care căutăm un corespondent. Pentru fiecare candidat, se calculează un scor de corelație care ia în calcul vecinătatea (fereastră de corelare care aici are dimensiunea de 3×3). Candidatul pentru care se obține cel mai bun scor este reținut ca pixel corespondent (punctul negru)

Ceea ce noi numim abuziv măsură de corelație poate de fapt cuantifica o similaritate sau, la fel de bine, o neasemănare. De aceea, în cele ce urmează vom considera că scorul „cel mai bun” corespunde maximului unei măsuri de similaritate sau minimului unei măsuri de neasemănare.

Candidatul din zona de căutare a cărei vecinătate se aseamănă cel mai mult cu cea a pixelului luat în considerare, adică cel care a obținut cel mai bun scor de corelație, este selecționat ca și corespondent. Această strategie poartă numele de strategie winner-take-all (wta – câștigătorul ia totul). Algoritmul 1 rezumă diferitele etape ale punerii în corespondență prin metoda clasică locală.

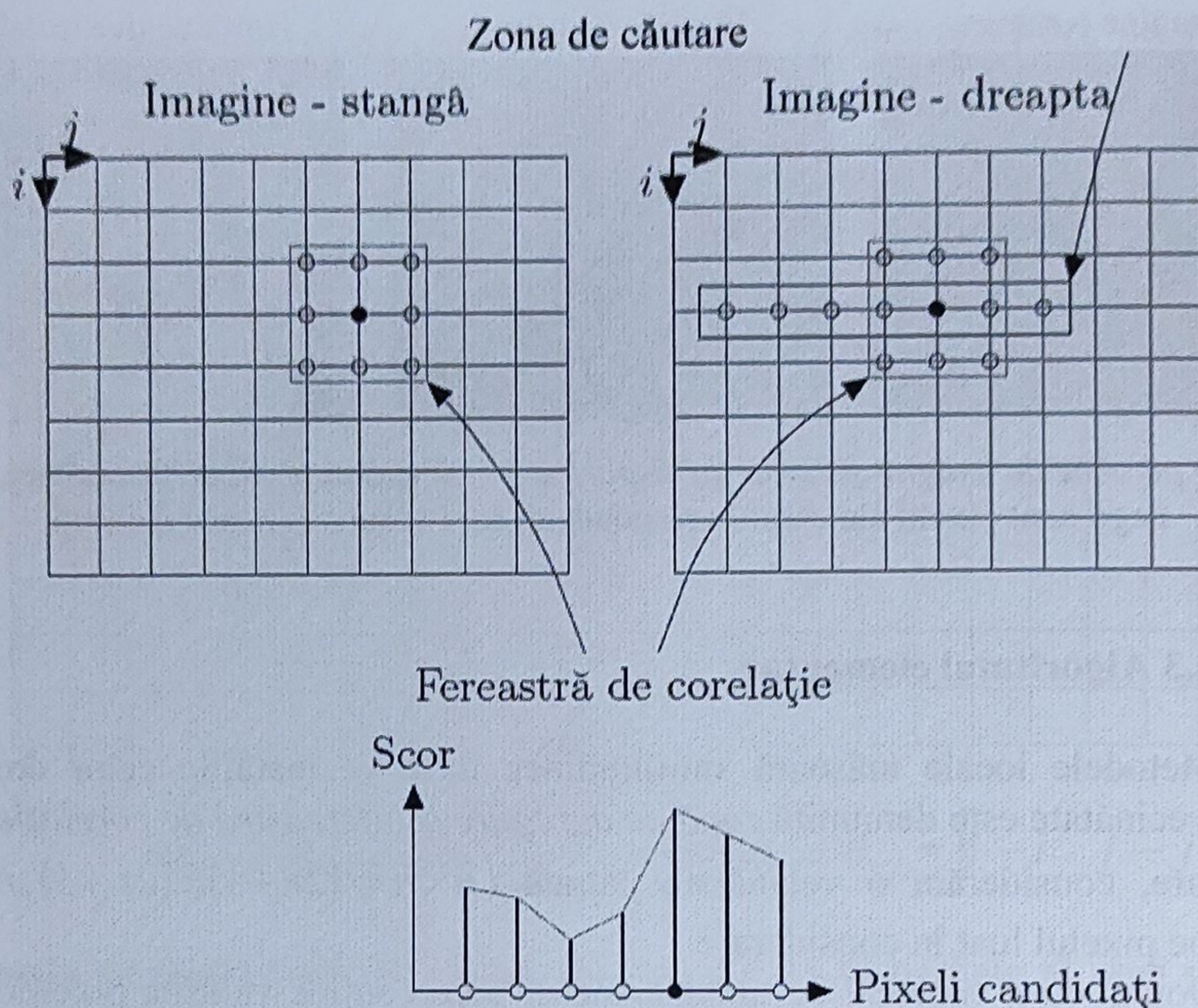


Fig. 9. Stabilirea corespondenței locale

Și mai formal, definim o măsură de corelație generică c prin:

$$c : N^{n^2}, N^{n^2} \rightarrow N$$

$$a, b \propto c(a, b).$$

(7)

Există erori frecvente datorate algoritmului elementar, în particular în zonele acoperite sau apropiate de discontinuitățile de profunzime (am văzut, în paragraful 1.3, că ipoteza expresiei (6) nu este întotdeauna verificată).

Pentru a limita aceste erori, este posibil să fie utilizate:

- **condiții:** care evita împerecherea unor pixeli care nu îndeplinesc anumite proprietăți;
- **măsuri de corelație adaptate la context:** anumite măsuri de corelație au fost concepute pentru a face față anumitor dificultăți, cum ar fi, de exemplu, apropierea de discontinuitățile de profunzime;
- **ferestre de corelație adaptate la context:** obiectivul este de a adapta morfologia ferestrei de corelație astfel ca să nu se ia în calcul decât partea din vecinătate care este cea mai pertinentă (totalitatea pixelilor aparținând vecinătăților pătrate a doi pixeli corespondenți nu se aseamănă neapărat în apropierea zonelor ascunse sau cu discontinuități de profunzime, conform paragrafului 1.3).

2.4 Utilizarea condițiilor

Distingem două tipuri de condiții: a priori și a posteriori.

Condiția a priori este utilizată pentru a limita riscul de a efectua o împerechere eronată prin reducerea zonei de căutare cu ajutorul geometriei epipolare: ne referim la condiția de epipolaritate.

Condițiile a posteriori sunt utilizate după stabilirea corespondențelor între pixeli, pentru eliminarea împerecherilor care nu sunt suficient de fiabile. Totuși, ele pot elimina în măsură egală unele perechi bune, reducând astfel densitatea rezultatului obținut. Condițiile de punere în corespondență sunt numeroase și nu vom prezenta aici decât pe cele mai des utilizate.

```

1:  $s_{\max} \leftarrow -\infty$ 
2: execută pentru fiecare pixel  $p_{i,j}^s$  din imaginea stânga
3:   execută pentru fiecare pixel candidat  $p_{i',j'}^d$  din zona de căutare din
      imaginea dreapta
4:      $s \leftarrow c(a,b)$ , scorul de corelație între valorile pixelilor din vecină-
      tățile lui  $p_{i,j}^s$  (stocate în vectorul  $a$ ) și  $p_{i',j'}^d$  (stocate în vectorul  $b$ )
5:     dacă  $s > s_{\max}$  atunci
6:        $s_{\max} \leftarrow s$ 
7:        $\hat{i}' \leftarrow i'$ 
8:        $\hat{j}' \leftarrow j'$ 
9:     sfârșit dacă
10:   sfârșit execută
11:    $m([i \ j]) \leftarrow [\hat{i}' \ \hat{j}']$ 
12: sfârșit execută
    
```

ALGORITMUL 1 – Algoritmul elementar de punere în corespondență clasică locală.

2.4.1 Condiția de epipolaritate

Așa cum am discutat deja în paragraful 2.1, modelul proiecției în perspectivă aduce o condiționare geometrică ce permite limitarea căutării corespondentului unui pixel la dreapta epipolară asociată acelui pixel. În practică, perechile de imagini suferă o rectificare epipolară care reduce zona de căutare corespunzătoare unui pixel situat pe linia i și coloana j în imaginea stânga la pixelii din imaginea dreapta care sunt situați la stânga coloanei j de pe aceeași linie i .

2.4.2 Condiția asupra scorului de corelație

Această condiționare, care este probabil cea mai des utilizată, constă în eliminarea perechilor pentru care scorul de corelație este inferior unui prag, adică nu sunt păstrate decât perechile $(p_{i,j}^s, p_{i',j'}^d)$ pentru care:

$$c(a, b) > t_c \quad (8)$$

unde a este vectorul valorilor pixelilor din vecinătatea lui $p_{i,j}^s$, b este vectorul valorilor pixelilor din vecinătatea lui $p_{i',j'}^d$, iar $t_c \in R$ este un prag.

2.4.3 Condiția de ordine

Ordinea pixelilor de-al lungul dreptei epipolare trebuie să fie aceeași cu cea a corespondenților lor de-a lungul dreptei epipolare conjugate (conform Fig. 10):

$$(m([i \ j_1]^T) = [i \ j_1]^T \wedge m([i \ j_2]^T) = [i \ j_2]^T \wedge j_1 < j_2) \Rightarrow j'_1 < j'_2 \quad (9)$$

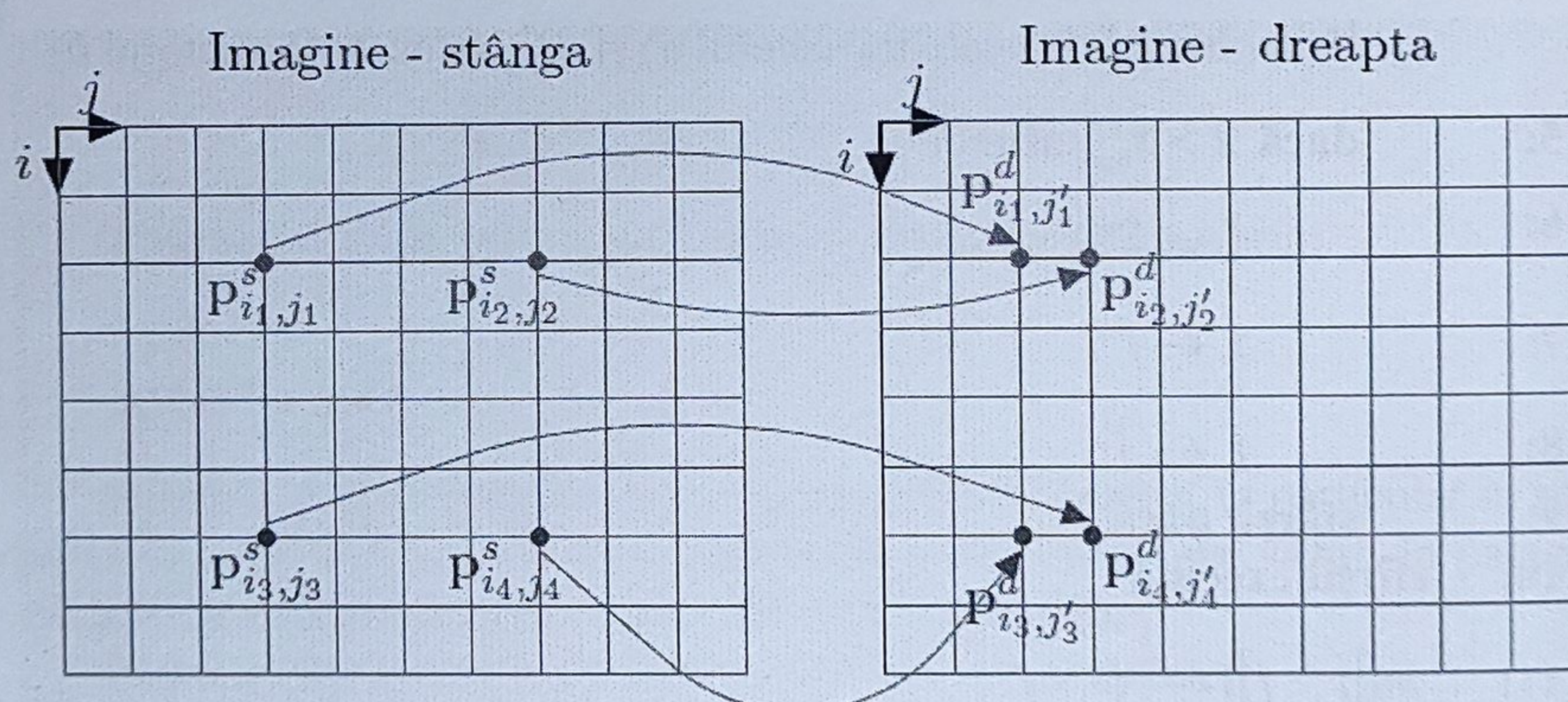


Fig. 10. Condiția de ordine – În partea de sus prezentăm un exemplu în configurație paralelă în care condiția de ordine este îndeplinită, în timp ce în partea de jos avem un exemplu în care aceasta nu este îndeplinită

2.4.4 Condiția de unicitate

Această condiție stipulează că un pixel dintr-una dintre imagini poate fi împerecheat cu cel mult un pixel din cealaltă imagine (conform Fig. 11). Condiția de unicitate este îndeplinită când:

$$\begin{aligned} (m^{s \rightarrow d}([i \ j_1]^T) = [i \ j_1]^T \Rightarrow \forall j_2 \neq j_1, m^{s \rightarrow d}([i \ j_2]^T) \neq [i \ j_1]^T) \\ \wedge (m^{d \rightarrow s}([i \ j'_1]^T) = [i \ j'_1]^T \Rightarrow \forall j'_2 \neq j'_1, m^{d \rightarrow s}([i \ j'_2]^T) \neq [i \ j'_1]^T) \end{aligned} \quad (10)$$

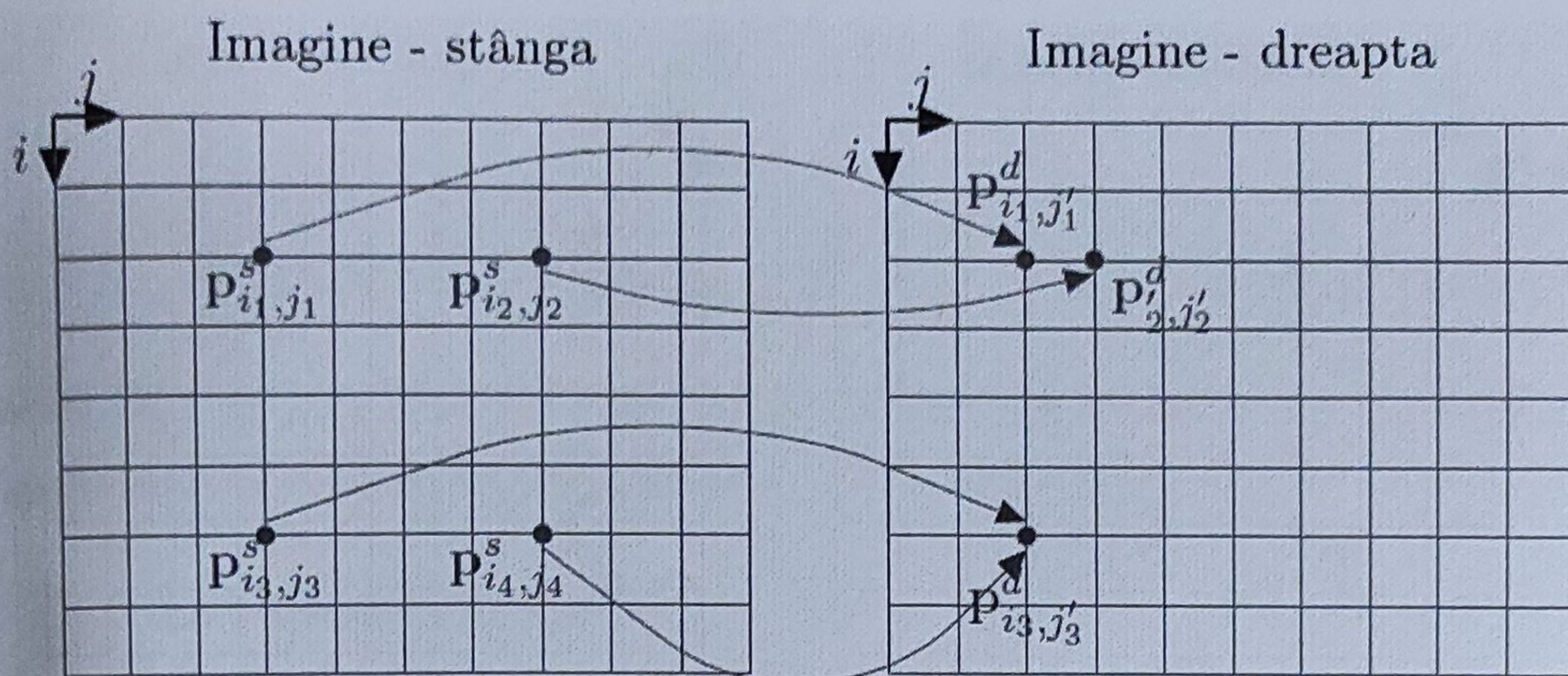


Fig. 11. *Condiția de unicitate* – În partea de sus prezentăm un exemplu în care condiția de unicitate este îndeplinită, în timp ce în partea de jos avem un exemplu în care aceasta nu este îndeplinită

2.4.5 Condiția de simetrie

Condiția de simetrie introdusă în [Fua, 1993]⁷, constă în satisfacerea condiției de regăsire a acelorași împerecheri indiferent care este imaginea de referință (conform Fig. 12):

$$m^{s \rightarrow d} \left(\begin{bmatrix} i & j \end{bmatrix}^T \right) = \begin{bmatrix} i & j' \end{bmatrix}^T \Leftrightarrow m^{d \rightarrow s} \left(\begin{bmatrix} i & j' \end{bmatrix}^T \right) = \begin{bmatrix} i & j \end{bmatrix}^T \quad (11)$$

Această condiție poate fi impusă prin efectuarea unei verificări bidirecționale, adică prin efectuarea unei puneri în corespondență a imaginii stânga față de imaginea dreapta, a unei puneri în corespondență a imaginii dreapta față de imaginea stânga și prin îndeplinirea faptului că regăsim bine aceleași împerecheri în ambele sensuri.

În Fig. 12 se ilustrează condiția de simetrie și de simetrie aproximativă. În partea de sus este prezentat un exemplu în care condiția de simetrie este îndeplinită, în timp ce în partea de jos este un exemplu în care aceasta nu este îndeplinită. În același timp, în acest exemplu, condiția de simetrie aproximativă este îndeplinită dacă luăm $t_{sf} = \varepsilon$.

⁷ Fua, Pascal, „A parallel stereo algorithm that produces dense depth maps and preserves image features”, International Journal of Machine Vision and Applications, 6:35–49, 1993.

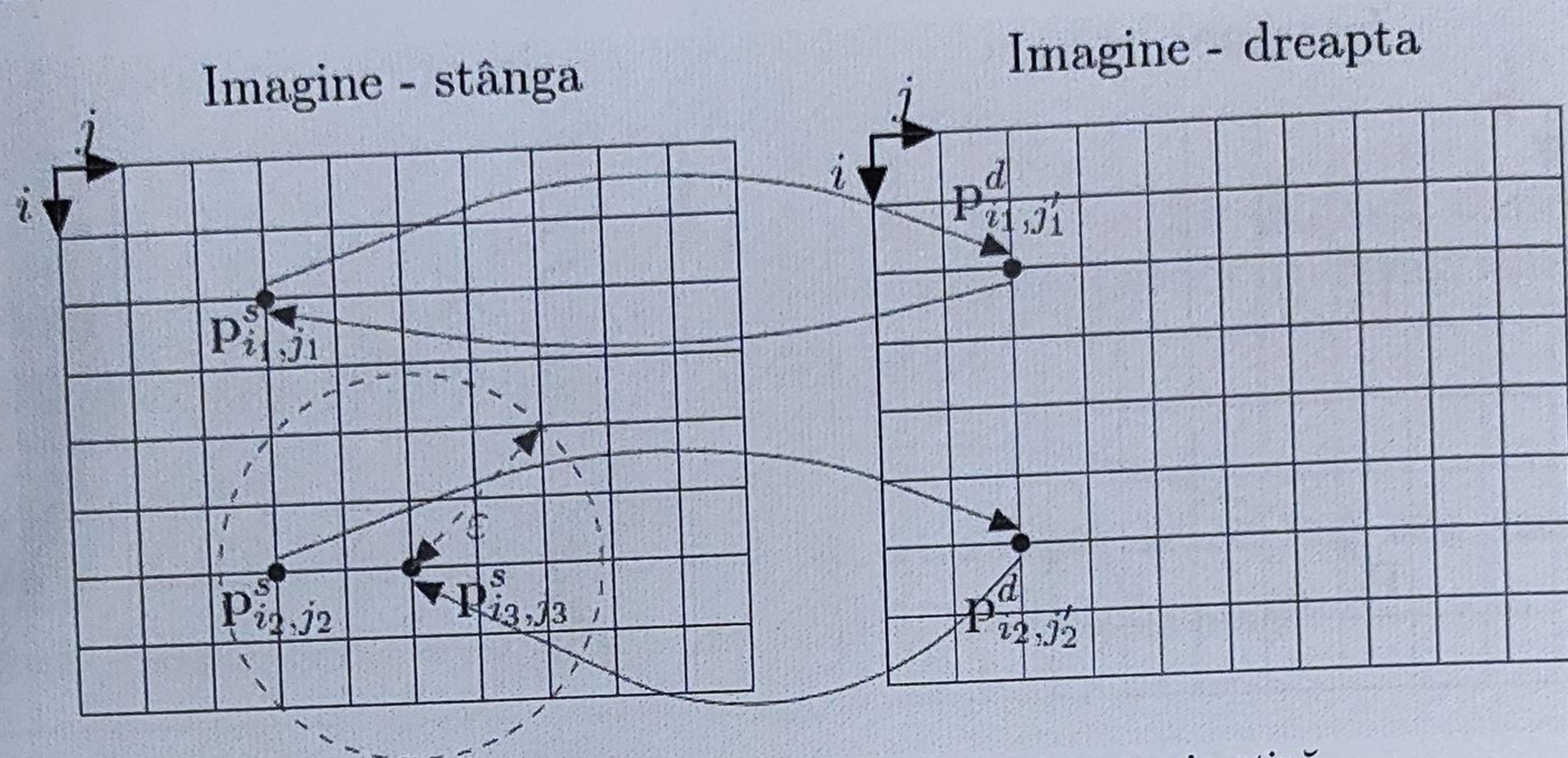


Fig. 12. Condiția de simetrie și de simetrie aproximativă

2.4.6 Condiția de simetrie aproximativă

Condiția precedentă este foarte puternică, astfel că îndeplinirea ei poate elimina multe dintre împerecherile corecte. O variantă mai tolerantă, pusă în operă printr-o verificare bidirecțională suplă, și care utilizează un prag $t_{sf} \in R$, este dată de:

$$m^{s \rightarrow d}([i \ j]^T) = [i \ j']^T \wedge m^{d \rightarrow s}([i \ j']^T) = [i \ j_2]^T \Rightarrow |j - j_2| < t_{sf} \quad (12)$$

2.4.7 Consistența aproximativă

Consistența aproximativă prezentată în [Gong, 2003]⁸ este o alternativă la condiția de simetrie. Dacă un punct al scenei se proiectează în $p_{i,j}^s$ și $p_{i,j}^d$, atunci corespondentul lui $p_{i,j}^s$ este $p_{i,j}^d$, mai puțin dacă el este ascuns. În acest caz, el nu poate fi acoperit decât de un punct mai apropiat de camera de luat vederi și având, deci, o disparitate mai mare, care se traduce prin:

$$\text{Dacă } j + d_{i,j} = j_2 \text{ atunci pentru toți } j' > j, j' + d_{i,j'} \neq j_2 \text{ și pentru toți } j'_2 < j_2, j'_2 + d_{i,j'_2} \neq j \quad (13)$$

2.4.8 Limita gradientului de disparitate

Această condiție impune că disparitățile nu variază brusc între doi pixeli vecini. Este definită de:

⁸ Gong, Minglun, Yang, Yee-Hong, „Fast stereo matching using reliability-based dynamic programming and consistency constraints”, IEEE International Conference on Computer Vision, Nisa, Franța, vol. 1, pp. 610–617, octombrie 2003.

$$m\left(\begin{bmatrix} i_1 & j_1 \end{bmatrix}^T\right) = \begin{bmatrix} i_1 & j_1 \end{bmatrix}^T \wedge m\left(\begin{bmatrix} i_2 & j_2 \end{bmatrix}^T\right) = \begin{bmatrix} i_2 & j_2 \end{bmatrix}^T \Rightarrow$$

$$\frac{\left\| \left(\begin{bmatrix} i_1 & j_1 \end{bmatrix}^T - \begin{bmatrix} i_2 & j_2 \end{bmatrix}^T \right) - \left(\begin{bmatrix} i_1 & j_1 \end{bmatrix}^T - \begin{bmatrix} i_2 & j_2 \end{bmatrix}^T \right) \right\|}{2 \left\| \left(\begin{bmatrix} i_1 & j_1 \end{bmatrix}^T - \begin{bmatrix} i_2 & j_2 \end{bmatrix}^T \right) + \left(\begin{bmatrix} i_1 & j_1 \end{bmatrix}^T - \begin{bmatrix} i_2 & j_2 \end{bmatrix}^T \right) \right\|} < t_g \quad (14)$$

unde p_{i_1, j_1}^s și p_{i_2, j_2}^s sunt doi pixeli care sunt proiecțiile a doi pixeli din scenă care aparțin unei aceleiași suprafețe și unde $t_g \in R$ este un prag.

2.4.9 Condiția de rang

Această condiție, prezentată în [Banks, 2001]⁹, compară valoarea pixelului central cu cele ale celorlalți pixeli din fereastra de corelație și impune ca vecinii ale căror valori sunt superioare celei a pixelului central să fie aceeași pentru cei doi pixeli corespondenți. În exemplul dat în Fig. 13, în prima linie sunt prezentate valorile conținute într-o fereastră de corelație de 3x3 în jurul pixelului considerat din stânga (a) și în jurul a doi candidați din dreapta (b și c). În cea de-a doua linie, sunt prezentate, pentru fiecare fereastră, diferențele între valoarea pixelului central și cele ale fiecărui vecin al său. Constatăm astfel că în exemplul dat, aceste diferențe pentru candidatul (b) au toate același semn ca și cele ale pixelului considerat (a): condiția de rang este respectată. În schimb, nu acesta este cazul pentru toate diferențele candidatului (c) (cu albine): condiția de rang nu este respectată. Această condiție poate fi scrisă:

$$m\left(\begin{bmatrix} i & j \end{bmatrix}^T\right) = \begin{bmatrix} i & j \end{bmatrix}^T \Rightarrow$$

$$\forall (k, l) \in \left[-\frac{n}{2}; \frac{n}{2}\right]^2, \left(I_{i+k, j+l}^s - I_{i, j}^s\right) \left(I_{i+k, j'+l}^d - I_{i, j'}^d\right) \quad (15)$$

	Stânga (a)			(b)			Dreapta (c)		
Nivele de gri	250	200	198	255	205	203	255	205	100
	100	104	110	105	109	115	105	109	105
	50	48	78	55	53	93	55	53	110
Diferență	146	96	94	146	96	94	146	96	-9
	-4	0	6	-4	0	6	-4	0	-4
	-54	-56	-26	-54	-56	-16	-54	-56	1

Fig. 13. Condiția de rang

⁹ Banks, Jasmine, Bennamoun, Mohammed, „Reliability analysis of the rank transform for stereo matching”, IEEE Transactions on Systems, Man, and Cybernetics, 31(6):870–880, 2001.

2.4.10 Condiția de fiabilitate

Aplicarea unui prag asupra unor măsuri de fiabilitate permite eliminarea împerecherilor mai puțin fiabile. Fie r o funcție care măsoară fiabilitatea pornind de la m scoruri de corelație între pixelul considerat $p_{i,j}^s$ și cei m candidați p_{i_k,j_k}^d din zona de căutare cu $k \in [0; m]$:

$$r: R^m \rightarrow R$$

$$\begin{bmatrix} M \\ c(a,b)_k \\ M \end{bmatrix} \alpha r \left(\begin{bmatrix} M \\ c(a,b)_k \\ M \end{bmatrix} \right) \quad (16)$$

Notăm $c = [K \quad c(a,b)_k \quad K]^T$. Condiția poate fi scrisă acum:

$$r(c) < t_f \quad (17)$$

unde $t_f \in R$ este un prag. Putem să distingem următoarele măsuri de fiabilitate (funcții r):

- *Măsură de ambiguitate* – măsoară distanța în pixeli dintre candidatul care a obținut cel mai bun scor de corelație și cel care are cel mai bun al doilea scor (conform Fig. 14). Fie:

$$amb(c) = |(c)_{1:m} - (c)_{2:m}| \quad (18)$$

unde $(c)_{1:m}$ și $(c)_{2:m}$ sunt primele două elemente ale vectorului c ale cărui elemente sunt triate în ordine descrescătoare. Măsura de ambiguitate nu este calculată decât dacă cel de-al doilea cel mai mare scor de corelație este apropiat de primul. Fie $d_{12} = |(c)_{1:m} - (c)_{2:m}|$. Măsura de ambiguitate este dată de:

$$AMB(c) = \begin{cases} amb(c) & \text{dacă } d_{12} < t_a \\ 0 & \text{dacă nu} \end{cases} \quad (19)$$

unde $t_a \in R$ este un prag.

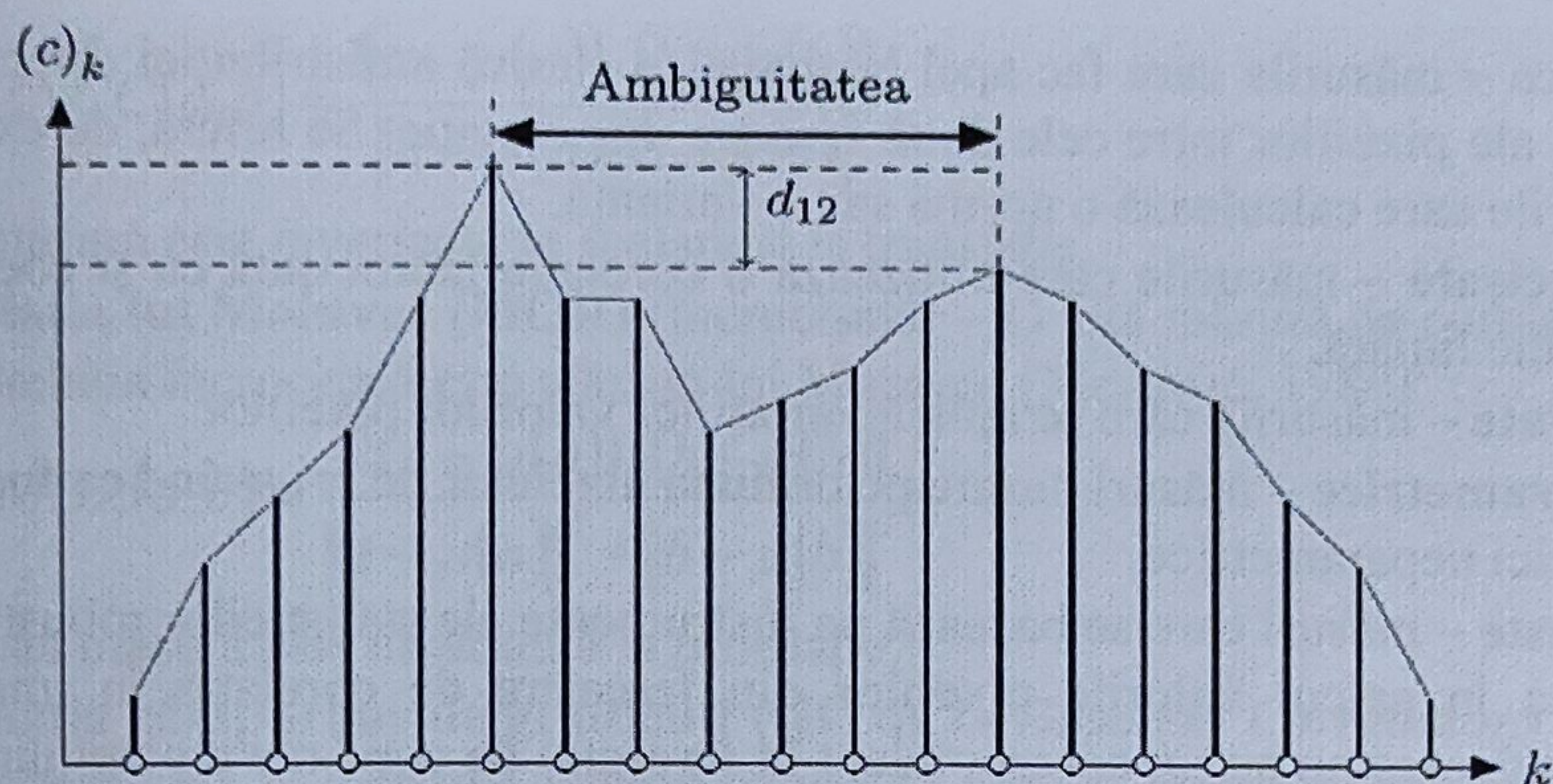


Fig. 14. *Măsura de ambiguitate* – Candidații din zona de căutare sunt marcați pe abscisă. Măsura de ambiguitate corespunde distanței (în pixeli) dintre cei doi candidați care au obținut cele mai bune două scoruri de corelație (în ordonată).

- *Măsură de imprecizie* – este dată de distanța dintre pixelii k_1 și k_2 care definesc intervalul în care scorurile sunt apropiate de scorul cel mai bun, adică sunt foarte mari (conform Fig. 15):

$$IMP(c) = |k_1 - k_2|, \text{ unde } k_1 \text{ și } k_2 \text{ sunt astfel ca } \forall k \in [k_1; k_2], c_k > t_i \quad (20)$$

unde $t_i \in R$ este un prag.

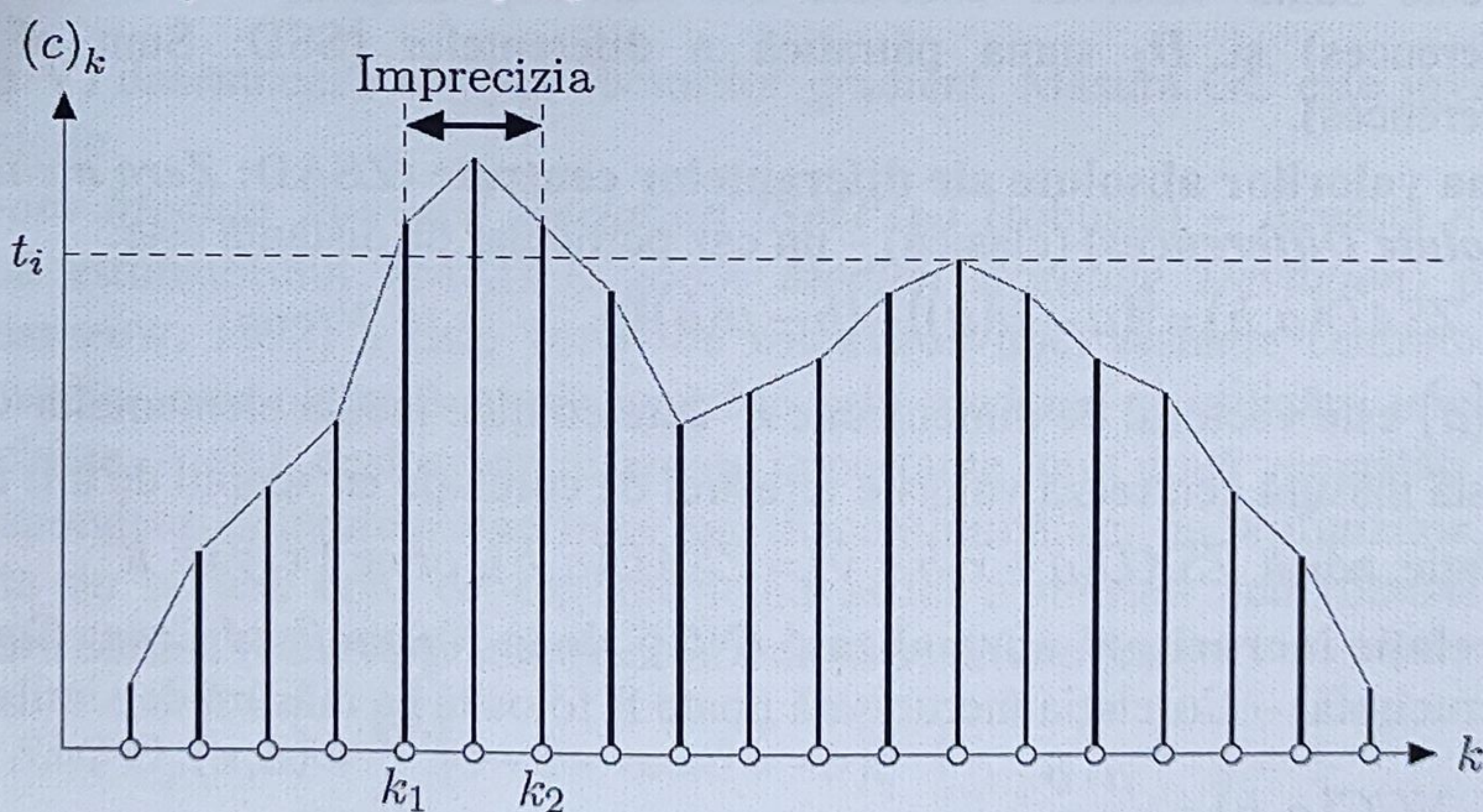


Fig. 15. *Măsura de imprecizie* – Candidații din zona de căutare sunt marcați pe abscisă, măsura de imprecizie corespunde numărului de pixeli pentru care scorul de corelație, în ordonată, este superior unui prag t_i .

2.5 Măsuri de corelație

Au fost propuse numeroase măsuri de corelație; ele pot fi clasificate în cinci familii [Chambon, 2011]¹⁰:

¹⁰ Chambon, Sylvie, Crouzil, Alain, „Similarity measures for image matching despite occlusions in stereo vision”, Pattern Recognition, 44:2063-2075, 2011.

- **Clasice** – măsurile care fac apel la statistica clasică a distribuției diferențelor de valori ale pixelilor între cele două ferestre de corelație. Se referă, de exemplu, la măsurile care calculează o normă sau o varianță.
- **Încrucișate** – măsurile care utilizează o corelație încrucișată ca și coeficient de corelație liniară.
- **Derivate** – măsurile care se aplică variațiilor valorilor pixelilor.
- **Neparametrice** – măsuri bazate pe ordinea nivelelor de gri având ca fundament statistici neparametrice.
- **Robuste** – măsuri care se bazează pe instrumente ale statisticilor robuste pentru a nu lua în calcul valorile pixelilor din fereastra de corelație în imaginea de referință care nu au un corespondent în fereastra de corelație a pixelului candidat din cealaltă imagine (cu precădere în cazul acoperirii sau discontinuității de profunzime).

Vom prezenta mai jos, pentru fiecare familie, măsurile de corelație care dau cele mai bune rezultate după evaluarea realizată în [Chambon, 2011]¹¹ precum și măsurile care vor fi utilizate în continuarea acestui capitol.

O taxonomie detaliată a măsurilor de corelație pentru stabilirea corespondențelor pixelilor poate fi, de asemenea, găsită în [Chambon, 2011]¹².

- **Norma L_P** (clasică) – se referă la o măsură de distanță:

$$D_P(a, b) = \|a - b\|_P^P. \quad (21)$$

- D_1 este suma valorilor absolute ale diferențelor (SAD: Sum of Absolute Differences) și D_2 suma pătratică a diferențelor (SSD: Sum of Squared Differences).

- **Suma valorilor absolute ale diferențelor centrate (ZSAD: Zero mean Sum of Absolute Differences)** (clasică) – un caz particular de distanță este:

$$ZSAD(a, b) = \|(a - \mu(a)) - (b - \mu(b))\|_1 \quad (22)$$

unde $\mu(a)$ este vectorul de dimensiune n^2 care conține media elementelor vectorului a . Această măsură centrează valorile ferestrei de corelație cu scopul de a fi invariantă la translație, adică $ZSAD(a + x, b + y) = ZSAD(a, b)$, orice $(x, y) \in R^2$.

- **Corelație încrucișată normalizată (NCC de la Normalized Cross Correlation)** (încrucișată) – Corelația încrucișată poate fi folosită ca măsură de similaritate:

$$NCC(a, b) = \frac{a \cdot b}{\|a\| \|b\|} \quad (23)$$

- unde \cdot reprezintă produsul scalar și unde valorile ferestrelor de corelație sunt normalizate cu scopul ca măsura să fie invariantă la scalare, adică $NCC(ax, by) = NCC(a, b)$, $\forall (x, y) \in R^{*2}$.

- **Corelație încrucișată centrată normalizată (ZNCC de la Zero mean Normalized Cross Correlation)** (încrucișată) – Există, de asemenea, o versiune centrată a NCC, coeficientul de corelație liniară:

¹¹ Idem.

¹² Idem.

$$ZNCC(a, b) = \frac{(a - \mu(a)) \cdot (b - \mu(b))}{\|a - \mu(a)\| \|b - \mu(b)\|} \quad (24)$$

Această măsură este invariantă la scalare și la translație.

- **Corelația lui Moravec (MOR)** (încrucișată) – O altă măsură de corelație bazată pe corelația încrucișată este măsura lui Moravec [Moravec, 1980]¹³:

$$MOR(a, b) = \frac{2(a - \mu(a)) \cdot (b - \mu(b))}{\|a - \mu(a)\|^2 + \|b - \mu(b)\|^2} \quad (25)$$

Această măsură este invariantă la translație.

- **Corelația câmpurilor de gradienti (GC de la Gradient Correlation)** (derivată) – O alternativă la măsurile clasice constă în a nu fi direct interesați de valorile pixelilor, ci de variația acestora prin intermediul primei lor derivate. Corelația gradientilor introdusă în [Crouzil, 1996]¹⁴ măsoară disimilitudinea vectorilor gradient conținuți în cele două vecinătăți și poate fi scrisă astfel:

$$GC(a, b) = \frac{\sum_{p=-\frac{n}{2}}^{\frac{n}{2}} \sum_{q=-\frac{n}{2}}^{\frac{n}{2}} \|\nabla I^s([i+p \ j+q]^T) - \nabla I^d([i+p \ j'+q]^T)\|}{\sum_{p=-\frac{n}{2}}^{\frac{n}{2}} \sum_{q=-\frac{n}{2}}^{\frac{n}{2}} \|\nabla I^s([i+p \ j+q]^T)\| + \|\nabla I^d([i+p \ j'+q]^T)\|} \quad (26)$$

unde ∇I desemnează câmpul vectorilor gradient. Măsura GC este invariantă la translație.

- **Smooth Median Powered Deviation (SMPD_p)** (robustă) – Această măsură se aplică estimatorului SMAD (*Smooth Median Absolute Deviation*) propus în [Rousseeuw, 1992]¹⁵. Este vorba de calcularea unei distanțe centrate robuste. Robustețea este obținută grație centrării prin mediană și utilizării unei distanțe trunchiate. Astfel, deși există diferențe importante între două vecinătăți de pixeli corespondenți datorate acoperirilor sau discontinuităților de profunzime, în cazul acesta nu se ține cont de ele, rezultând astfel o măsură mai robustă decât o măsură clasică (care va ține cont de aceste mari diferențe):

$$SMPD_p(a, b) = \sum_{k=0}^{h-1} \left(|a - b - med(a - b)|^p \right)_{kn^2-1} \quad (27)$$

¹³ Moravec, Hans.P., „Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover”, Teză de doctorat, Universitatea Stanford, Stanford, California, S.U.A., septembrie 1980.

¹⁴ Crouzil, Alain, Massip-Pailhes, Louis, Castan, Serge, „A new correlation criterion based on gradient fields similarity”, International Conference on Pattern Recognition, vol. 1, pp. 632-636, Viena, Austria, august 1996.

¹⁵ Rousseeuw, Peter J., Croux, Christophe, „L_p-statistical analysis and related methods”, în Y. Dodge, editor, Explicit Scale Estimators with High Breakdown point, pp. 77-92, Elsevier, Amsterdam, Olanda, 1992.

unde $med(a-b)$ este mediana diferențelor valorilor celor două ferestre de corelație și $k : n^2 - 1$ este al elementul de ordin k al vectorului de abateri de la mediană triat în ordine crescătoare. În general se ia $h = n^2 / 2$.

- **CENSUS** (familia neparametrică) – Această măsură, introdusă în [Zabih, 1994]¹⁶, ține cont de diferențele de valori între pixelul studiat și vecinii săi. Fie R^a și R^b două șiruri de biți definiți în următoarea manieră:

$$R^a = \bigotimes_{k \in [0; n^2]} \begin{cases} 1 & \text{dacă } a_{\frac{n^2}{2}} > a_k \\ 0 & \text{dacă } nu \end{cases} \quad R^b = \bigotimes_{k \in [0; n^2]} \begin{cases} 1 & \text{dacă } b_{\frac{n^2}{2}} > b_k \\ 0 & \text{dacă } nu \end{cases} \quad (28)$$

unde \bigotimes este operatorul de concatenare și k parcurge vecinătatea (de dimensiune n^2) a fiecărui pixel din fereastra de corelație. Scorul este dat de:

$$CENSUS(a, b) = d_{Ham}(R^a, R^b) \quad (29)$$

unde d_{Ham} este distanța Hamming, adică numărul de biți care diferă între cele două șiruri. Această măsură dă, de asemenea, rezultate bune în zonele apropiate de acoperiri și de discontinuități de profunzime în [Chambon, 2011]¹⁷. Această măsură evaluează respectarea condiției de rang (conform paragrafului 2.4.9). Fig. 16 prezintă disparitățile perechii de imagini *Conuri* (vezi Fig. 8) obținute cu măsura CENSUS.



Fig. 16. Harta disparităților perechii de imagini *Conuri* obținute cu măsura CENSUS (vecinătate 9×9)

¹⁶ Zabih, Ramin, Woodfill, John, „Non-parametric local transforms for computing visual correspondence”, European Conference on Computer Vision, pp. 151-158, Stockholm, Suedia, mai 1994.

¹⁷ Chambon, Sylvie, Crouzil, Alain, „Similarity measures for image matching despite occlusions in stereo vision”, Pattern Recognition, 44:2063-2075, 2011.

2.6 Vecinătăți adaptabile

O dimensiune prea mică a ferestrei de corelație nu ține cont de atâta informație încât să fie suficient de discriminantă și să facă față unor anumite dificultăți, în particular în zonele omogene sau constituite din texturi regulate. În plus, măsura riscă să fie prea sensibilă la mici variații puțin semnificative (zgomot). Pe de altă parte, o fereastră de corelație prea mare riscă să conțină pixeli care nu au aceeași disparitate ca și cea a pixelului considerat. Astfel, pentru doi pixeli corespondenți, vom obține vecinătăți care nu sunt întotdeauna similare (conform paragrafului 1.3). Măsura de corelație utilizată ține cont de aceste diferențe și, dacă aceasta nu este adaptată, dă scoruri scăzute pentru doi corespondenți și poate implica erori de împerechere. De aceea au fost propuse diferite tehnici care permit adaptarea mărimii sau formei ferestrelor de corelație pentru a nu ține cont decât de pixelii din vecinătate care corespund unul altuia. Se pot distinge următoarele metode:

- **deformare legată a ferestrelor de corelație:** ferestrele de corelație au aceeași formă în imaginea de referință și în cealaltă imagine;
- **deformare independentă a ferestrelor de corelație:** ferestrele de corelație nu au aceeași formă în imaginea de referință și în cealaltă imagine;
- **ponderarea pixelilor din fereastră de corelație:** o pondere este atribuită fiecărui pixel din fereastră și intervine în calculul scorului de corelație.

2.6.1 Deformare legată a ferestrelor de corelație

- **minimizarea unui criteriu de incertitudine** – În [Kanade, 1994]¹⁸, autorii au propus un algoritm iterativ care estimează și actualizează pentru fiecare pixel valoarea de disparitate găsită, adaptând la fiecare iterație dimensiunile ferestrei de corelație utilizată. Aceste dimensiuni sunt determinate prin testarea mai multor dimensiuni posibile și prin selecționarea acelor care minimizează un criteriu de incertitudine aplicat disparității estimate în cadrul iterației precedente. Calculul acestui criteriu de incertitudine este bazat pe un model statistic care ține cont de distribuția diferențelor dintre disparitățile estimate pentru pixelul considerat și cele estimate pentru ceilalți pixeli din fereastră de corelație, precum și distribuția variațiilor valorilor pixelilor din fereastră de corelație a candidatului din cealaltă imagine. Astfel, fereastră care minimizează acest criteriu este utilizată pentru actualizarea valorii disparității estimate a pixelului considerat și procesul se repetă până la convergență.
- **ferestre multiple fixe** – În [Bobick, 1999]¹⁹, sunt utilizate nouă ferestre de corelație diferite. Prima este centrată pe pixelul considerat, celelalte opt sunt excentrice. Scorul final este dat de fereastră care dă scorul de similitudine cel

¹⁸ Kanade, Takeo, Okutomi, Masatoshi, „A stereo matching algorithm with an adaptive window: Theory and experiment”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(9): pp. 920-932, 1994.

¹⁹ Bobick, Aaron F., Intille, Stephen S., „Large occlusion stereo”, International Journal of Computer Vision, 33(3): pp. 181-200, 1999.

mai ridicat. Se presupune că este vorba despre scorul dat de fereastra care ține cont cel mai mult de pixelii vecini de aceeași disparitate ca și pixelul considerat. Această tehnică este cunoscută sub numele de *shiftable windows*.

- **ferestre multiple cu încrucișare** – În [Lotti, 1994]²⁰, autorii presupun că rupturile de profunzime coincid cu contururile din imagine. O primă etapă constă în detectarea contururilor din imagine cu ajutorul unui filtru Canny-Deriche și a aplicării unui prag cu histeresis. Patru ferestre pătrate de dimensiuni 1×1 sunt centrate mai apoi pe pixelul considerat. Fiecare fereastră este asociată uneia din cele patru direcții nord-vest, nord-est, sud-est, sud-vest și este lărgită progresiv cât timp ea nu se suprapune pe contururile prezente în zona studiată. Scorul de corelație final este calculat considerând reunirea celor patru ferestre.
- **ferestre multiple adiacente** – În [Hirschmuller, 2002]²¹, o fereastră de corelație este centrată pe pixelul considerat și apoi alte ferestre sunt plasate în jur, suprapunându-se peste ea. Fiecare fereastră dă un scor de corelație și scorul final este compus din suma celor mai ridicate scoruri.

2.6.2 Deformare independentă a ferestrelor de corelație

Metodele prezentate mai sus descriu, pentru fiecare cuplu de pixeli corespondenți, ferestre de corelație de formă identică în imaginea de referință și în cealaltă imagine fără să ia în calcul eventualele deformări ale obiectului considerat între cele două imagini, stânga și dreapta. În acest caz, este de preferat să se modifice forma ferestrei de corelație în cealaltă imagine astfel ca să ia în calcul pixelii corespondenți ai pixelilor incluși în fereastra de corelație ce aparține imaginii de referință.

Deformarea ferestrei este abordată printr-o transformare afină a cărei parametri pot fi estimați prin optimizare sau pornind de la distribuția valorilor pixelilor în vecinătatea pixelilor corespondenți.

- **Estimarea prin optimizare** – Această abordare este utilizată în [Devernay, 1994]²² [Garcia, 2001]²³ pentru stabilirea corespondențelor prin corelație fină, caz în care parametrii transformării locale a ferestrei (deformare și translație) sunt integrați măsurii de corelație. Stabilirea corespondențelor prin corelație fină este utilizată, de asemenea, în [Harvent, 2010]²⁴. O idee similară este utilizată în

²⁰ Lotti, Jean-Luc, Giraudon, Gerard, „Correlation algorithm with adaptive window for aerial image in stereo vision”, în European Symposium on Satellite Remote Sensing, pp. 2315-2325, Roma, septembrie 1994.

²¹ Hirschmuller, Heiko, Innocent, Peter R., Garibaldi, Jon, „Real-time correlation-based stereo vision with reduced border errors”, International Journal of Computer Vision, 47(1-3): pp. 229-246, 2002.

²² Devernay, Frederic, Faugeras, Olivier, „Computing differential properties of 3-D shapes from stereoscopic images without 3-D models”, raport de cercetare RR-2304, Institut National de Recherche en Informatique et en Automatique, INRIA, 1994.

²³ Garcia, Dorian, „Mesures de formes et de champs de déplacements tridimensionnels par stereo-correlation d'images”, teză de doctorat, Institut National Polytechnique de Toulouse, decembrie 2001.

²⁴ Harvent, Jacques, „Mesures de formes par correlation multi-images: application a l'inspection de pieces aeronautiques a l'aide d'un systeme multi-cameras”, teză de doctorat, Universitatea din Toulouse, noiembrie 2010.

[Gruen, 1985]²⁵. Acești parametri trebuie apoi să maximizeze măsura de corelație adaptată care a fost aleasă. Ei sunt estimați printr-o tehnică de optimizare (conform paragrafului 4.3) care depinde de măsura de corelație aleasă și care necesită o estimare inițială a parametrilor. Parametrii translației sunt inițial estimați pornind de la o hartă de disparitate calculată în prealabil cu o metodă clasică. Ceilalți parametri sunt considerați inițial nuli. Vectorii de disparitate rezultă în urma optimizării tuturor acestor parametri.

- **Pornind de la distribuția valorilor pixelilor în vecinătățile corespondențelor** – În [Kannala, 2007]²⁶ autorii estimează în vecinătatea fiecărui pixel corespondent parametrii unei elipse, în funcție de distribuția locală a valorii pixelilor. Transformarea afină locală corespunde apoi transformării între cele două elipse ale celor doi pixeli corespondenți. Această metodă este utilizată pentru stabilirea corespondențelor prin propagare, așa cum vom vedea în paragraful 3.2.

2.6.3 Ponderarea pixelilor din fereastra de corelație

O pondere este atribuită fiecărui pixel din fereastra de corelație centrată pe pixelul considerat. Această pondere reprezintă influența fiecărui pixel vecin în calculul scorului de corelație.

- **Mascarea zonelor de acoperire** – În [Kostkrova, 2002]²⁷, se face o primă punere în corespondență cu ferestre de corelație centrate pe pixelii considerați. După aceea se realizează o a doua punere în corespondență, caz în care se construiește o fereastră de corelație nouă pentru fiecare pixel în funcție de primele disparități estimate, astfel încât numai pixelii din vecinătate care au aceeași disparitate estimată cu cea a pixelului studiat, să fie luați în seamă la calculul scorului de corelație. Avem aici de fapt o ponderare binară.
- **Ponderare fotometrică și spațială** – În [Yoon, 2005]²⁸, ponderea atribuită fiecărui pixel din fereastra de corelație este proporțională cu probabilitatea ca pixelul vecin considerat să aibă aceeași disparitate ca și pixelul central. Această probabilitate este estimată ținând cont de diferența de culoare dintre cei doi pixeli, precum și de distanța care îi separă.

În [Gong, 2007]²⁹, o versiune simplificată a acestei ultime tehnici este evaluată și comparată cu alte metode de agregare folosind ferestre pătrate și ferestre

²⁵ Gruen, Armin W., „Adaptive least squares correlation: a powerful image matching technique”, South African Journal of Photogrammetry, Remote Sensing and Cartography, vol. 3, nr. 14, pp. 175-187, 1985.

²⁶ Kannala, Juho, Brandt, Sami S., „Quasi-dense wide baseline matching using mec propagation”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1-8, Statele Unite, iunie 2007.

²⁷ Kostkrova, Jana, Sara, Radim, „Disparity component matching revisited”, Raport tehnic CTU-CMP-2002-08, Universitatea Tehnică din Praga, 2002.

²⁸ Yoon, Kuk-Jin, Kweon, In-So, „Locally adaptive support-weight approach for visual correspondence search”, IEEE Computer Society conference on Computer Vision and pattern recognition, vol. 2, pp. 924-931, San Diego, California, Statele Unite, iunie 2005.

²⁹ Gong, Minglun, Yang, Ruigang, Wang, Liang, Gong, Mingwei, „A performance study on different cost aggregation approaches used in real-time stereo matching”, International Journal of Computer Vision, vol. 2, nr. 75, pp. 283-296, 2007.

multiple diferite. Metoda cu ponderare este cea care furnizează cele mai puține împerecheri eronate.

Remarcă: se poate considera că tehnicile cu ponderare se aseamănă cu metodele bazate pe instrumentele statisticilor robuste. Cititorul interesat poate găsi detalii suplimentare în [Chambon, 2011]³⁰.

2.7 Utilizarea culorii

Pentru a utiliza toate informațiile furnizate de culoare în cadrul imaginilor au fost propuse strategii de generalizare a metodei locale de stabilire a corespondențelor. În [Chambon, 2005]³¹, au fost luate în considerare nouă spații de reprezentare a culorii și au fost explorate patru variante:

- **Fuziunea scorurilor** – pentru fiecare pixel candidat se calculează câte un scor de corelație cu fiecare componentă a spațiului de culoare. Apoi scorurile care rezultă sunt fuzionate printr-un operator pentru a obține scorul final.
- **Fuziunea hărților de disparitate** – se efectuează o punere în corespondență prin utilizarea în mod independent a fiecărei componente de culoare a imaginii. Hărțile de disparitate astfel obținute sunt fuzionate mai departe cu ajutorul unui simplu mecanism de vot. Dacă hărțile de disparitate furnizează pixeli corespondenți diferiți, atunci cel pentru care rezultă cel mai bun scor este selecționat.
- **Generalizarea directă a măsurii de corelație** – se referă la modificarea formulelor cu care se calculează măsurile de corelație astfel ca ele să ia în considerare mărimi de intrare vectoriale (culorile) în loc de mărimi de intrare scalare (nivelele de gri). Pentru aceasta, fiecare operator de bază (produs scalar, norme, distanțe, relații de ordine ...) utilizat în calculul acestor măsuri a fost generalizat pentru culori.
- **Analiza componentelor principale** – ultima variantă constă în realizarea unei puneri în corespondență de tip clasic aplicată primei componente care rezultă din analiza componentelor principale a imaginilor color.

Rezultatele experimentale prezentate în [Chambon, 2005]³² arată că utilizarea culorii duce la ameliorarea rezultatelor punerii în corespondență a pixelilor, dar cu prețul unei creșteri a timpului de calcul.

În particular, numărul de împerecheri corecte este mai ridicat și verificarea bidirecțională (conform paragrafului 2.4.5) furnizează mai puține rezultate fals negative față de utilizarea nivelelor de gri, mai ales în apropierea marginilor obiectelor. Pe de altă parte, nu este clară alegerea celui mai bun spațiu de culoare, precum și a celei mai bune strategii de utilizare a culorii, deoarece acestea sunt legate de măsura de corelație folosită. Cu toate acestea, în linii mari se pare că fuziunea scorurilor de corelație cu operatorul de maximizare și spațiul de culoare RGB se constituie într-un bun compromis care este convenabil în numeroase situații.

³⁰ Chambon, Sylvie, Crouzil, Alain, „*Similarity measures for image matching despite occlusions in stereo vision*”, Pattern Recognition, 44:2063-2075, 2011.

³¹ Chambon, Sylvie, Crouzil, Alain, „*Colour correlation-based matching*”, International Journal of Robotics and Automation, 20(2):78-87, 2005.

³² *Idem*.

3. Stabilirea corespondențelor între pixeli prin metoda creșterii regiunilor

Am arătat în paragraful 2.4.1 că zona de căutare a unui pixel corespondent poate fi redusă la o singură dreaptă prin utilizarea geometriei epipolare. Această zonă de căutare poate fi redusă și mai mult dacă se ia în considerare ipoteza că, aproape în toate cazurile, pixelii vecini au disparități apropiate. Pe această ipoteză se bazează metodele care utilizează propagarea. Este vorba despre algoritmi iterativi care se aplică unui set de perechi credibile de puncte de start numite germenii („seeds” în engleză).

În cele ce urmează vom utiliza următorii termeni:

- **Germen** – o pereche credibilă, adică o pereche de pixeli căruia îi acordăm o încredere sporită. Setul de germenii aparținând unei perechi de imagini la momentul t se notează cu G_t (setul inițial fiind notat cu G_0).
- **Pixelul stânga al unui germen** – se referă la pixelul situat în imaginea din stânga (de referință în cazul nostru). Dacă pixelul $p_{i,j}^s$ este pixelul de referință al unui germen, atunci îl vom nota cu $g_{i,j}^s$. Notăm cu $\mathcal{R}_{G_t}^s$ setul de pixeli stânga ai germenilor G_t .
- **Pixelul dreapta al unui germen** – se referă la pixelul corespondent în imaginea din dreapta a pixelului stânga al unui germen. Dacă pixelul $p_{i',j'}^d$ este corespondentul unui pixel de referință al unui germen, atunci îl vom nota cu $g_{i',j'}^d$. Notăm cu $\mathcal{R}_{G_t}^d$ setul de pixeli dreapta ai germenilor G_t .
- **Funcția de corespondență a germenilor** – Este definită astfel:

$$g^{s \leftrightarrow d} : \mathcal{R}_{G_t}^s \rightarrow \mathcal{R}_{G_t}^d$$

$$g_{i,j}^s \rightarrow g_{i',j'}^d$$
(30)
- **Fereastră de vecinătate** – Se referă la fereastra centrată pe pixelul stânga al germenului considerat care include pixelii vecini pentru care vor fi căutați corespondenții în vecinătatea pixelului dreapta al germenului. Pentru simplificare, vom considera o vecinătate pătrată de dimensiuni $n_v \times n_v$ (conform Fig. 19).
- **Zona de căutare** – În cazul general, se referă la zona de căutare a candidaților din vecinătatea pixelului dreapta al germenului considerat. Pentru simplificare, vom considera o zonă de căutare pătrată de dimensiuni n_z^2 . În cazul în care imaginile au fost rectificate, zona de căutare poate fi redusă și mai mult până la zona marcată cu gri în Fig. 19, constituită doar din $n_z/2 - 1$ pixeli.

3.1 Selecționarea germenilor inițiali

Stabilirea corespondențelor prin propagare este o alternativă interesantă la metodele locale clasice. Reducerea zonei de căutare a pixelilor corespondenți la

vecinătățile germenilor nu numai că reduce volumul de calcul, dar în aceeași măsură reduce riscul selecționării unui corespondent greșit, totul ducând la diminuarea ambiguităților. Cu toate acestea, este necesară o atenție deosebită la selecția germenilor inițiali.

Distingem două familii de metode de selectare a germenilor:

- **Împerecherea punctelor de interes** – Punctele de interes sunt acei pixeli speciali din imagine care au proprietăți care permit o împerechere mai credibilă, adică pentru care riscul de a comite erori de punere în corespondență este limitat. Există diferiți detectori de puncte de interes. Un detector „bun” este repetabil [Schmid, 2000]³³, adică trebuie să fie capabil să detecteze pixelii care sunt proiecții ale acelorași elemente ale scenei în imagini diferite. În felul acesta, la punerea în corespondență a unui punct de interes, candidații sunt reduși la un sub-set mic de pixeli. Aceasta prezintă mai puține ambiguități în alegerea corespondentului decât dacă s-ar fi testat, de exemplu, toți pixelii de pe dreapta epipolară corespondentă. Această abordare o găsim utilizată în [Otto, 1989]³⁴ [Lhuillier, 2002]³⁵ [Kannala, 2007]³⁶.
- **Perechi de interes** – Nu există un criteriu de selecție prealabilă a pixelilor. Se aplică o metodă locală clasică de punere în corespondență tuturor pixelilor din imagine (sau numai unui subset de pixeli). Zona de căutare a candidaților este de aceea mai mare decât în cazul limitării la punctele de interes. Totuși, se utilizează un set de condiții tari pentru a elimina împerecherile care par mai puțin credibile. Perechile rămase, a priori considerate mai sigure, constituie setul inițial de germenii. Cu scopul de a ține sub control repartitia germenilor pe toată suprafața imaginilor, în [Chen, 1999]³⁷ se utilizează o tehnică de *bucketing*, care constă în împărțirea imaginii în blocuri și căutarea germenilor în cadrul fiecărui bloc.

Proprietatea de repetabilitate a detectorilor de puncte de interes asigură că pixelii care sunt proiecțiile unei aceleiași entități din scenă sunt detectați în mai multe imagini ale aceleiași scene. În felul acesta, detectarea punctelor de interes permite limitarea căutărilor la un set de câțiva candidați printre care se găsesc foarte probabil pixelii corespondenți. Această abordare este deci interesantă pentru selecția germenilor inițiali care se derulează în două etape: detecția urmată de împerecherea punctelor de interes.

³³ Schmid, Cordelia, Mohr, Roger, Bauckhage, Christian, „*Evaluation of interest point detectors*”, International Journal of Computer Vision, 37(2):151-172, 2000.

³⁴ Otto, G. Paul, Chau, Tony K. W., „*Region-growing*» algorithm for matching of terrain images”, Image and Vision Computing, 7(2):83-94, 1989.

³⁵ Lhuillier, Maxime, Quan, Long, „*Match propagation for image-based modeling and rendering*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(8): 1140-1146, 2002.

³⁶ Kannala, Juho, Brandt, Sami S., „*Quasi-dense wide baseline matching using meci propagation*”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1-8, Statele Unite, iunie 2007.

³⁷ Chen, Qian, Medioni, Gérard, „*A volumetric stereo matching method: Application to image-based modeling*”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pp. 1029-1034, Fort Collins, SUA, iunie 1999.

3.1.1 Detectarea punctelor de interes

Prima etapă permite detectarea pixelilor din imagine care sunt *a priori* mai susceptibili decât alții să fie împerecheați corect. Un punct de interes este un punct particular din imagine care posedă caracteristici interesante pentru o aplicație dată. De exemplu, punctele de interes sunt utilizate în cadrul urmăririi obiectelor într-o secvență de imagini [Parisot, 2009]³⁸ sau în cadrul indexării [Mikolajczyk, 2002]³⁹ și recunoașterii imaginilor [Jurie, 2004]⁴⁰. În cazul nostru, ne referim la un punct ce poate fi pus în corespondență într-o manieră credibilă, adică înconjurat de pixeli cu valori care prezintă variații ce îi caracterizează ușor. Acest lucru permite descrierea detaliată a unei astfel de vecinătăți și furnizarea de informații ajutătoare procesului de punere în corespondență a pixelilor.

O proprietate interesantă a detectorilor de puncte de interes este repetabilitatea. Criteriul repetabilității [Schmid, 2000]⁴¹ este raportul dintre numărul de puncte de interes repetate și numărul de puncte de interes detectate. În cazul nostru, un punct de interes poate fi numit „repetat” dacă corespondentul său este de asemenea detectat ca și punct de interes. O marjă de eroare de ε pixeli este în general tolerată. În Fig. 17, P_1 și P_2 sunt două puncte ale scenei și proiecțiile lui P_1 au fost detectate ca și puncte de interes în imaginea stânga și în imaginea dreapta (încercuite cu negru). Punctul din imaginea stânga care este proiecția lui P_1 este repetat în imaginea dreapta, contrar proiecției lui P_2 . Cercurile de rază ε denotă o marjă de toleranță asupra erorii de detectare în considerarea unui punct ca fiind repetat sau nu. De exemplu, dacă un punct de interes ar fi fost prezent în cercul marcat cu circumferință continuă, atunci am fi considerat proiecția lui P_2 ca repetată. Detectorii de puncte de interes sunt utilizați pentru proprietățile lor de robustețe în fața diferitelor transformări fotometrice și geometrice (punctele de interes sunt repetate în pofida prezenței acestor transformări între cele două imagini). Aceste transformări geometrice nu apar în mod obișnuit în contextul în care discutăm (cele două camere de luat vederi nu sunt poziționate prea îndepărtat una de cealaltă). Problemele dominante cu care suntem confrunțați sunt cele de obstrucție și discontinuitate a profunzimii.

În [Tuytelaars, 2008]⁴², diferiții detectori de puncte de interes sunt clasificați după tipul de punct de interes căutat:

³⁸ Parisot, Pascaline, „Suivi d'objets dans les séquences d'images de scènes déformables: de l'importance des points d'intérêt et du maillage 2D”, Teză de doctorat, Universitatea din Toulouse, ianuarie 2009.

³⁹ Mikolajczyk, Krystian, „Detection of local features invariant to affine transformations. Application to matching and recognition”, Teză de doctorat, Institut National Polytechnique de Grenoble, iunie 2002.

⁴⁰ Jurie, Frédéric, Schmid, Cordelia, „Scale-invariant shape features for recognition of object categorie”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 90–96, Washington, DC, S.U.A., iunie 2004.

⁴¹ Schmid, Cordelia, Mohr, Roger, Bauckhage, Christian, „Evaluation of interest point detectors”, International Journal of Computer Vision, 37(2):151–172, 2000.

⁴² Tuytelaars, Tinne, Mikolajczyk, Krystian, „Local invariant feature detectors: a survey”, Foundations and Trends® in Computer Graphics and Vision, 3(3):177–280, 2008.

- **Colțuri** – Un colț corespunde intersecției mai multor muchii ale unui obiect (joncțiuni „L”, „T” sau „Y”). Un colț detectat nu corespunde în mod necesar unui colț al unui obiect din scenă. Totuși, putem presupune că punctele de interes situate pe colțuri (adică la marginile obiectelor) riscă să fie aproape de rupturile de profunzime sau de obstrucții.
- **Blob-uri** – O definiție formală a blob-ului este dată în [Lindeberg, 1993]⁴³ și putem să o rezumăm în felul următor: se referă la o zonă din imagine care înconjoară un punct (de interes) în care valorile pixelilor se detașează net de valorile pixelilor din vecinătate. De exemplu, este un blob o zonă dintr-o imagine care se prezintă ca o pată luminoasă situată într-o zonă întunecată. În plus față de centrul blob-ului, un detector de acest tip trebuie în același timp să furnizeze caracteristica mărimii zonei luate în considerare.

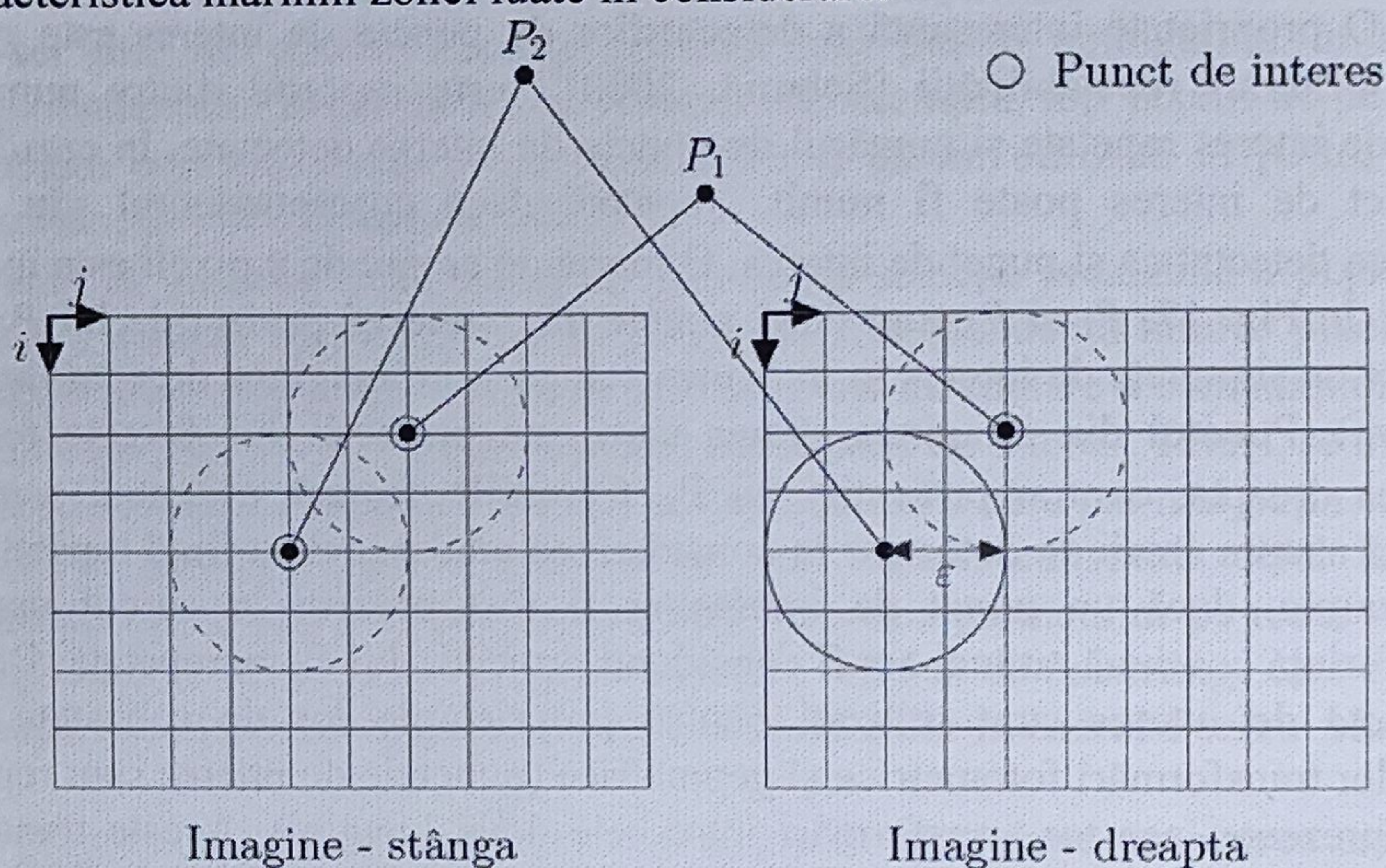


Fig. 17. Repetabilitatea unui detector de puncte de interes

În [Schmid, 2000]⁴⁴, diferiții detectori de puncte de interes sunt clasificate după tehnica de detecție:

- **Detectori bazați pe contururi** – Aceste metode detectează colțurile la nivelul intersecțiilor contururilor sau chiar la nivelul curburilor foarte accentuate ale contururilor.
- **Detectori bazați pe modele parametrice** – Aceste metode estimează parametrii modelului unui colț astfel ca el să se modifice pentru a se suprapune precis pe imagine la nivelul punctului luat în considerare.

⁴³ Lindeberg, Tony, „Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention”, International Journal of Computer Vision, 1(3):1074–1130, 1993.

⁴⁴ Schmid, Cordelia, Mohr, Roger, Bauckhage, Christian, „Evaluation of interest point detectors”, International Journal of Computer Vision, 37(2):151-172, 2000.

- **Detectori bazați pe intensitate** – Aceste metode calculează pentru fiecare pixel o „valoare de interes”, numită răspuns, care ia în considerare valorile pixelilor din jur.

Detectorii care se bazează pe intensități urmează în general următoarele etape:

1. **Calculează răspunsul pentru fiecare pixel** – Acest răspuns depinde de valorile pixelilor din vecinătatea pixelului considerat. Se pot utiliza în calcul derivatele, criteriile morfologice sau chiar entropia distribuției valorilor pixelilor. Dimensiunile vecinătăților luate în considerare sunt legate de noțiunea de *scală*. Dacă aceste dimensiuni sunt fixe, se spune că *detectorul este cu scală fixă*. În cazul în care aceste dimensiuni sunt variabile, se spune că *detectorul este multi-scală*. Notăm cu r funcția care dă răspunsul:

$$r : N^2 \rightarrow R^n$$

$$\begin{bmatrix} i \\ j \end{bmatrix} \alpha r \left(\begin{bmatrix} i \\ j \end{bmatrix} \right). \quad (31)$$

Această funcție poate da o singură valoare de răspuns ($n = 1$) în cazul detectorilor cu scală fixă sau poate da mai multe valori ($n > 1$) în cazul detectorilor multi-scală (câte o valoare pentru fiecare scală).

2. **Selecția extremelor locale ale răspunsului** – Sunt localizate extremele răspunsului. Aceste poziții constituie puncte de interes potențiale.
3. **Post-procesare** – Se referă la tehnicile care au în vedere eliminarea răspunsurilor false, cum ar fi aplicarea unui prag asupra valorii răspunsului, suprimarea candidaților situați pe contururi sau care nu respectă anumite condiții geometrice. De asemenea se poate referi la tehnicile care permit o poziționare la nivelul sub-pixelilor punctelor de interes, în care caz funcția care dă răspunsul este definită pe R^2 și are valori în R^2 .

Scala corespunde nivelului de detaliu la care se face observația. Cu cât este mai mică vecinătatea luată în considerare, cu atât se face o focalizare pe detalii (depinde de nivelul de zgomot); și invers, cu o vecinătate mai extinsă, se iau în considerare structurile mari cu riscul de a nu lua în considerare anumite detalii. Dacă un detector examinează mai multe scale, răspunsul și extremele sale sunt calculate într-un „spațiu scalat”. Se referă la o reprezentare multi-rezoluție a unei imagini la care, cu cât micșorăm scala, cu atât apar detalii noi, conservând însă pe cele existente la scale mai mari. În [Gales, 2011]⁴⁵ putem găsi o prezentare detaliată a mai multor detectori de puncte de interes, printre care detectorii cunoscuți de genul Harris [Harris, 1988]⁴⁶, SIFT [Lowe, 1999]⁴⁷, SURF [Bay, 2006]⁴⁸, SUSAN [Smith, 1997]⁴⁹, FAST [Rosten, 2006]⁵⁰, și variantele lor.

⁴⁵ Gales, Guillaume, „Mise en correspondance de pixels pour la stéréovision binoculaire par propagation d'appariements de points d'intérêt et sondage de régions”, teză de doctorat, Universitatea din Toulouse, iulie 2011.

⁴⁶ Harris, Chris, Stephens, Mike, „A combined corner and edge detector”, în Alvey Vision Conference, pag. 147–151, Manchester, UK, ianuarie 1988.

⁴⁷ Lowe, David G., „Object recognition from local scale-invariant features”, în IEEE International Conference on Computer Vision, vol. 2, pp. 1150–1157, Corfu, Grecia, septembrie 1999.

3.1.2 Stabilirea corespondențelor punctelor de interes

Cea de-a doua etapă permite punerea în corespondență a punctelor de interes. Un descriptor permite caracterizarea unui punct de interes. Ne referim la un vector care permite „descrierea” punctului. Acest vector poate fi compus, de exemplu, din nivelele de gri din vecinătatea punctului. Alți descriptori sunt concepuți pentru a fi invariante la transformările geometrice (rotație, schimbarea de scală, schimbarea perspectivei). O măsură de tip distanță permite apoi compararea descriptorilor între ei.

Punerea în corespondență a punctelor de interes este o problemă în sine în domeniul analizei imaginilor pentru că ea este utilizată în aplicații diferite, cum ar fi recunoașterea obiectelor, indexarea imaginilor sau urmărirea obiectelor în secvențe de imagini. Există numeroși descriptori a căror studiu comparativ poate fi găsit în [Mikolajczyk, 2005]⁵¹.

În cazul stereoviziunii în care cele două imagini sunt luate din poziții apropiate (*small-baseline*), nu există transformări geometrice importante între cele două imagini. În schimb, suntem confrunțați cu probleme de discontinuitate a profunzimii. În consecință, chiar dacă nu sunt invariante la transformările geometrice, se pot utiliza (conform paragrafului 2.5) metodele de împerechere prin măsurarea corelației robuste a nivelelor de gri (sau de culoare).

Cu toate acestea, punerea în corespondență a punctelor de interes diferă de o punere în corespondență densă la nivelul listei candidaților de examinat. De fapt, presupunem că un pixel corespondent al unui punct de interes din imaginea de referință este, de asemenea, un punct de interes al celeilalte imagini (pentru că punctele de interes generate de un detector „bun” sunt repetate). Rezultatele prezentate în [Schmid, 2000]⁵² au arătat că punctele de interes pot fi subiectul unei probleme de localizare. Aceasta se traduce prin faptul că un pixel corespondent al unui punct de interes din imaginea stânga poate fi un punct de interes în imaginea dreapta, dar, de asemenea și unul dintre vecinii săi. O punere în corespondență a punctelor de interes din imaginea de referință cu punctele de interes și vecinii lor din cealaltă imagine este, deci, o variantă mai potrivită decât punerea în corespondență a punctelor de interes doar cu punctele de interes.

Mai mult, ne așteptăm să găsim germenii care sunt fiabili, adică a căror împerechere să fie corectă, ceea ce, în mod evident, nu poate fi cunoscut „apriori”. Astfel, nu mai pare inutilă verificarea condițiilor asupra împerecherilor rezultate prin calcul. Au fost propuse numeroase condiții pentru a elimina împerecherile de puncte de interes mai puțin credibile, în special pornind de la geometria epipolară (conform

⁴⁸ Bay, Herbert, Ess, Andreas, Tuytelaars, Tinne, Gool, Luc Van, „SURF: Speeded up robust features”, *Computer Vision and Image Understanding*, 110(3):346–359, 2006.

⁴⁹ Smith, Stephen M., Brady, J. Michael, „SUSAN – a new approach to low level image processing”, *International Journal of Computer Vision*, 23(1):45–78, 1997.

⁵⁰ Rosten, Edward, Drummond, Tom, „Machine learning for high-speed corner detection”, în *European Conference on Computer Vision*, vol. 1, pag. 430–443, Graz, Austria, mai 2006.

⁵¹ Mikolajczyk, Krystian, Schmid, Cordelia, „A performance evaluation of local descriptors”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

⁵² Schmid, Cordelia, Mohr, Roger, Bauckhage, Christian, „Evaluation of interest point detectors”, *International Journal of Computer Vision*, 37(2):151–172, 2000.

paragrafului 2.4.1). În afară de asta, se pot aplica, de asemenea, condițiile prezentate în paragraful 2.4 pentru punerea în corespondență locală a pixelilor. Un studiu privind utilizarea acestor condiții diferite pentru punerea în corespondență a punctelor de interes poate fi găsit în [Vincent, 2001]⁵³[Zhou, 2002]⁵⁴[Sur, 2010]⁵⁵.

3.1.3 Obținerea unui set bun de germeni inițiali

Germenii inițiali sunt constituiți dintr-un set de împerecheri care servesc ca punct de plecare pentru punerea în corespondență prin propagare, este deci necesar ca să prezinte următoarele calități:

- **Fiabilitate** – Germenii inițiali trebuie să fie corecți, adică împerecherile realizate să fie reale, cu scopul de a evita propagarea prea multor erori la debutul procesului. În practică, nu se poate ști *a priori* dacă împerecherile rezultate în urma calculului sunt corecte.
- **Împrăștiere** – Este important ca germenii să fie „judicios” plasați în imagini, adică într-o manieră care să permită, prin propagare, găsirea de împerecheri peste tot în imagini.



Fig. 18. *Germenii inițiali obținuți pentru perechea de imagini Conuri*

În [Gales, 2011]⁵⁶, se prezintă o evaluare a metodelor de selecție a germenilor inițiali pentru împerecherea punctelor de interes. Parametrii diferitelor metode de împerechere a punctelor de interes care permit satisfacerea pe cât posibilă a compromisului dintre fiabilitatea și împrăștierea germenilor au fost determinați prin

⁵³ Vincent, Étienne, Laganière, Robert, „Matching feature points in stereo pairs: A comparative study of some matching strategies”, Machine Graphics & Vision, 10(3):237–259, 2001.

⁵⁴ Zhou, Ji, Shi, Jiaoying, „A robust algorithm for feature point matching”, Computer and Graphics, 26(3):429–436, 2002.

⁵⁵ Sur, Frédéric, Noury, Nicolas, Berger, Marie-Odile, „Determining point correspondences between two views under geometric constraint and photometric consistency”, Raport de cercetare RR-7246, Institut National de Recherche en Informatique et en Automatique, INRIA, 2010.

⁵⁶ Gales, Guillaume, „Mise en correspondance de pixels pour la stéréovision binoculaire par propagation d'appariements de points d'intérêt et sondage de régions”, teză de doctorat, Universitatea din Toulouse, iulie 2011.

optimizare. Rezultatele obținute arată că detectorul FAST sau detectorul SUSAN, asociate cu măsura de corelație CENSUS (sau măsura GC, mai puțin costisitoare din punctul de vedere al timpului de calcul), sunt un compromis bun între fiabilitate și împrăștiere. Apoi se propune o combinație de seturi de germeni complementari rezultați în urma aplicării unor metode diferite și având scopul de a obține un nou set care să se constituie într-un compromis și mai bun.

Acest set este apoi completat cu împerecheri de interes în scopul de a ameliora împrăștierea germenilor. Pentru aceasta, se realizează o segmentare fină a imaginii în regiuni de culoare omogenă prin metoda *mean-shift* [Comaniciu, 1997]⁵⁷. Pentru fiecare regiune care nu conține germeni, se realizează o căutare exhaustivă a corespondențelor printr-o abordare locală clasică, dar cu măsura de corelație robustă SMPD₂. Împerecherile astfel obținute care nu respectă condiția de simetrie sunt eliminați. În sfârșit, se efectuează o ultimă etapă de validare prin realizarea unei segmentări mai grosiere și prin eliminarea, înăuntrul fiecărei regiuni, a împerecherilor asociate disparităților foarte diferite de cele ale altor împerecheri ale regiunii (conform [Gales, 2011]⁵⁸ pentru detalii asupra acestei metode). Fig. 18 prezintă germenii inițiali obținuți pentru perechea de imagini *Conuri* (Fig. 8.).

3.2 Propagarea simultană și propagarea secvențială

Principiul propagării poate fi enunțat astfel. Pentru fiecare pixel $p_{i+k,j+l}^s$, $(k,l) \in [-n_v/2; n_v/2]^2$, vecin al pixelului din stânga $g_{i,j}^s$ a germenului considerat, candidații testați sunt cei din zona de căutare poziționată în jurul vecinului pixelului dreapta al acelui germen: $p_{i'+k+q,j'+l+r}^d$, $(q,r) \in [-n_z/2; n_z/2]^2$ (conform Fig. 19). Noile împerecheri care se stabilesc servesc apoi ca noi germeni și procesul este repetat atât timp cât sunt găsite noi disparități.

În figura 19, perechea $(g_{i,j}^s, g_{i',j'}^d)$ este un germen. Ferestrele de vecinătate sunt marcate cu chenar negru. Pentru pixelii vecini ai lui $g_{i,j}^s$ din imaginea stânga (marcați cu cerușe), căutarea pixelilor corespondenți se face în vecinătatea vecinilor pixelului corespondent al germenului luat în considerare. Mai precis, dacă luăm ca exemplu pixelul $p_{i+1,j+1}^s$, zona lui de căutare (fereastră marcată cu chenar gri) este constituită din pixelii $(q,r) \in [-n_z/2; n_z/2] \times [-n_z/2; n_z/2]$. Având în vedere că imaginile au suferit o rectificare epipolară, zona de căutare este redusă la sub-setul marcat cu gri.

⁵⁷ Comaniciu, Dorin, Meer, Peter, „Robust analysis of feature spaces: color image segmentation”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 750–755, San Juan, Porto Rico, iunie 1997.

⁵⁸ Gales, Guillaume, „Mise en correspondance de pixels pour la stéréovision binoculaire par propagation d'appariements de points d'intérêt et sondage de régions”, teză de doctorat, Universitatea din Toulouse, iulie 2011.

Se disting două abordări ale propagării în funcție de maniera în care sunt utilizați germenii:

- **Abordarea simultană** – La fiecare iterație t , toți germenii din setul G_t sunt luați în considerare simultan (conform Fig. 20). Aceasta implică că disparitatea calculată pentru un pixel poate fi indusă de mai mulți germenii. În acest caz, cea mai bună potrivire este păstrată, adică cea care dă cel mai bun scor de corelație și care îndeplinește și ansamblu de condiții impuse. Această abordare este propusă în [Chen, 1999]⁵⁹. În figura 20, pixelii pentru care a fost deja găsit un corespondent sunt marcați cu un punct negru. Germenii luați în considerare sunt punctele negre încercuite. Pixelii albi încercuiți sunt aceia pentru care se face căutarea unui corespondent în cadrul iterației curente. În abordarea simultană, toți germenii sunt luați în considerare în timpul aceleiași iterații. În acest exemplu, al iterației t , se face căutarea corespondențelor pentru 16 pixeli (cerculețele albe), și sunt găsiți, de exemplu, doar 14 (germenii iterației $t + 1$).

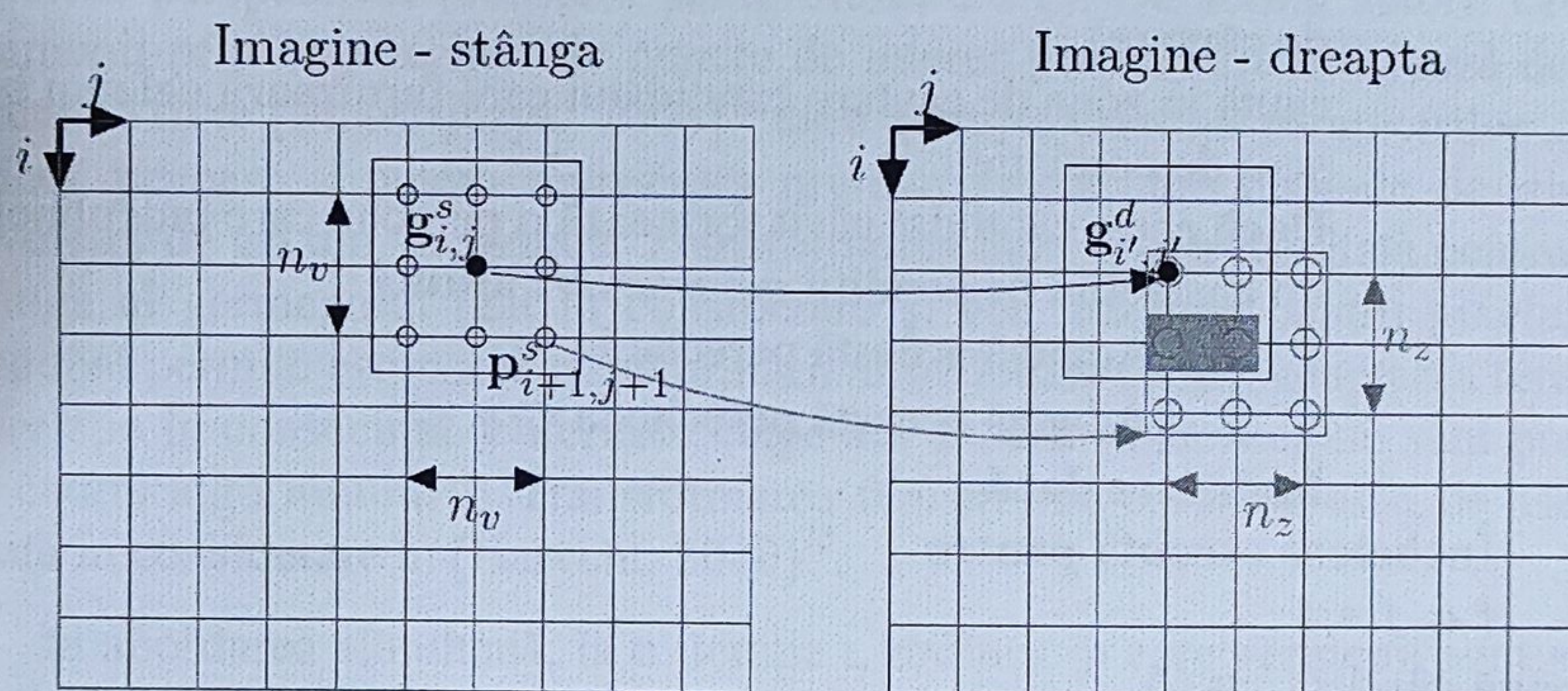


Fig. 19. Principiul propagării: Perechea $(g_{i,j}^s, g_{i',j'}^d)$ este un germen.

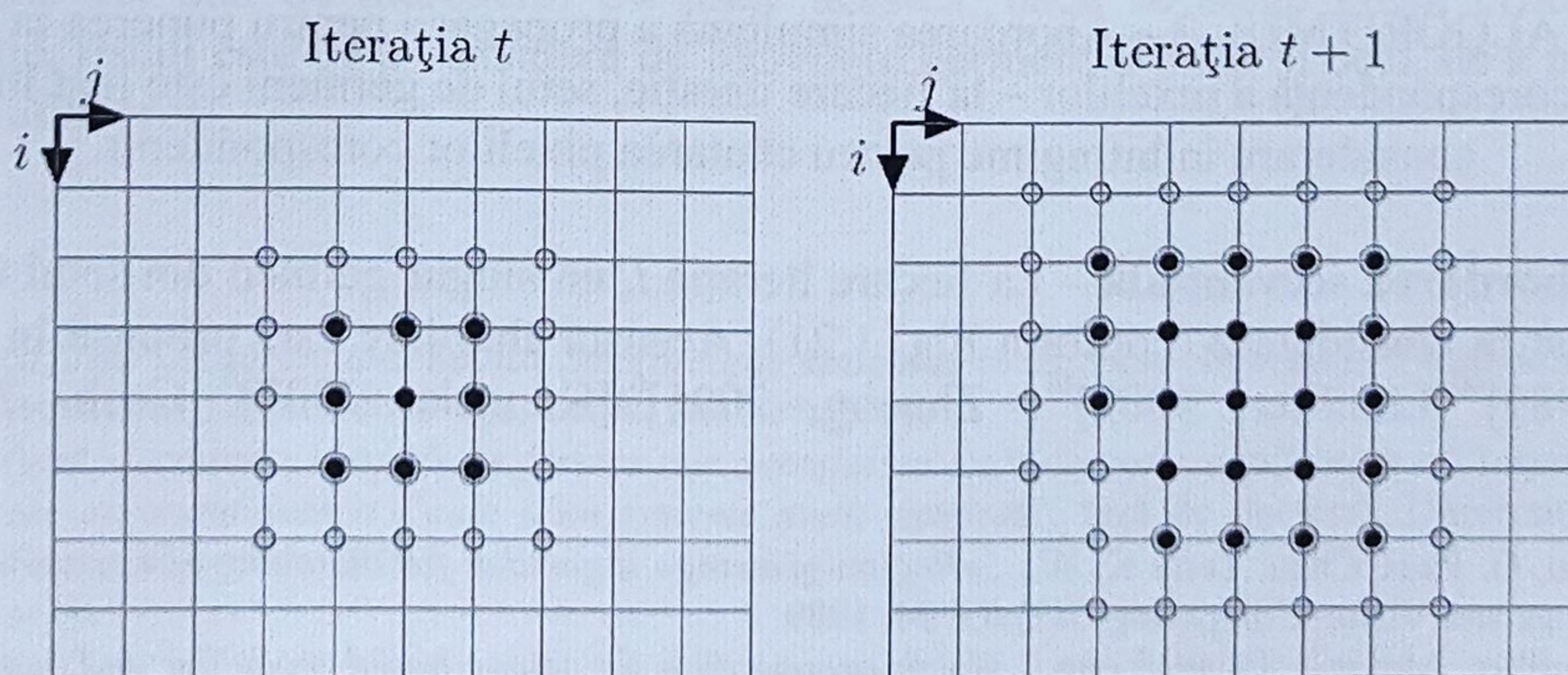


Fig. 20. Abordarea simultană a propagării

⁵⁹ Chen, Qian, Medioni, Gérard, „A volumetric stereo matching method: Application to image-based modeling”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pp 1029–1034, Fort Collins, SUA, iunie 1999.

Diferitele etape ale acestei abordări sunt prezentate prin algoritmul 2.

```

1:  $G_t$ : setul de germeni la iterația  $t$ 
2:  $\mathcal{M}$ : setul de împerecheri rezultate prin calcul
3:  $t \leftarrow 0$ 
4:  $G_t \leftarrow$  setul inițial de germeni
5:  $\mathcal{M} \leftarrow G_t$ 
6: Repetă
7:    $G_{t+1} \leftarrow \emptyset$ 
8:   Execută pentru toți pixelii din imaginea stânga care nu sunt germeni
9:     construiește zona de căutare, pornind de la toți germenii vecini și
       făcând o reuniune a fiecărei zone de căutare dată de fiecare din
       germenii vecini (conform Fig. 19 pentru construcția unei zone
       de căutare)
10:    caută în zona de căutare candidatul care furnizează cel mai bun
       scor de corelație
11:    Dacă acest candidat găsit formează o pereche care îndeplinește
       un ansamblu de condiții a posteriori atunci
12:      Adaugă această pereche la  $\mathcal{M}$ 
13:      Adaugă această pereche la  $G_{t+1}$ 
14:    Încheiere dacă
15:  Încheiere execută pentru
16:     $t \leftarrow t + 1$ 
17: până când  $G_{t+1} = \emptyset$ 
18: Returnează  $\mathcal{M}$ 
    
```

ALGORITMUL 2 – Abordarea simultană a propagării pentru punerea în corespondență a pixelilor – la fiecare iterație, setul de germeni este luat în considerare în întregime pentru căutarea pixelilor corespondenți.

- **Abordarea secvențială** – La fiecare iterație t , un singur germen din setul G_t este luat în considerare (conform Fig. 1.21). Această abordare este propusă în [Otto, 1989]⁶⁰[Lhuillier, 2002]⁶¹[Zhuang, 2005]⁶²[Kannala, 2007]⁶³[Cech, 2009]⁶⁴

⁶⁰ Otto, G. Paul, Chau, Tony K. W., „Region-growing» algorithm for matching of terrain images”, Image and Vision Computing, 7(2):83–94, 1989.

⁶¹ Lhuillier, Maxime, Quan, Long, „Match propagation for image-based modeling and rendering”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(8): 1140–1146, 2002.

⁶² Zhuang, Yong, Ma, Ma, Jun, Wei, Liu, Liu, „Progressive matching based on segmentation for 3D reconstruction”, în International Conference on Computer and Information Technology, pp. 575–579, Shanghai, China, septembrie 2005.

⁶³ Kannala, Juho, Brandt, Sami S., „Quasi-dense wide baseline matching using meci propagation”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1-8, Statele Unite, iunie 2007.

[Harvent, 2010]⁶⁵. Acel germen este selecționat pe baza unui criteriu predefinit și împerecherile găsite nu mai sunt puse sub semnul întrebării după aceea. Această strategie permite algoritmului să propage mai puține erori atunci când există un germen eronat, pentru că acesta nu este *a priori* ales cu prioritate de criteriul de selecție, care servește la tratarea în primul rând a germenilor „mai buni”. În [Lhuillier, 2002]⁶⁶ [Cech, 2009]⁶⁷, autorii selecționează germenii la fiecare iterație după scorul de corelație cel mai ridicat. În [Otto, 1989]⁶⁸, cea mai mare valoare proprie a matricei de covarianță obținută pornind de la disparitățile estimate este utilizată drept criteriu de selecție a germenilor. Conform autorilor, această valoare este reprezentativă pentru ordinul de mărime a erorii comise (metoda lor se aplică la imagini de terenuri cu puține discontinuități de profunzime). În [Kannala, 2007]⁶⁹, autorii reiau modelul din [Lhuillier, 2002]⁷⁰ și îl extind la cuplurile de imagini *wide-baseline*, adică preluate din poziții relative îndepărtate, ținând cont de transformările afine locale ale valorilor pixelilor datorate perspectivei (conform paragrafului 2.6.2). De aceea, autorii calculează germenii prin împerecheri de puncte de interes pentru care utilizează detectorul Harris-Affine [Mikolajczyk, 2004]⁷¹ care permite calculul parametrilor „regiunii afine” pe care o putem reprezenta printr-o elipsă care depinde de distribuția nivelelor de gri din vecinătate. Transformarea între elipsa rezultată pentru pixelul stânga al germenului luat în considerare și cea rezultată pentru pixelul corespondent permite estimarea transformării afine locale și deci adaptarea ferestrei de corelație în consecință. Diferitele etape ale abordării secvențiale sunt prezentate în cadrul algoritmului 3. Transformarea ferestrei de corelație este, de asemenea, luată în considerare în [Harvent, 2010]⁷².

În abordarea simultană, la o iterație t , ordinea în care germenii sunt selecționați nu influențează rezultatul. Totuși, dacă se efectuează o împerechere eronată, acea eroare riscă să se propage în cadrul iterațiilor următoare. Abordarea secvențială permite limitarea propagării erorilor prin tratarea cu prioritate a germenilor care sunt *a priori* cei mai de încredere. Cu toate acestea, dacă se alege greșit criteriul de selecție, riscul este de a considera de încredere germenii eronați și deci, de a propaga mai multe erori.

⁶⁴ Cech, Jan, „*Accurate and Robust Stereoscopic Matching in Efficient Algorithms*”, teză de doctorat, Universitatea Tehnică din Pragă, Cehia, februarie 2009.

⁶⁵ Harvent, Jacques, „*Mesures de formes par correlation multi-images: application a l'inspection de pieces aeronautiques a l'aide d'un systeme multi-cameras*”, teză de doctorat, Universitatea din Toulouse, noiembrie 2010.

⁶⁶ *Idem* 60.

⁶⁷ *Idem* 63.

⁶⁸ Otto, G. Paul, Chau, Tony K. W., „*«Region-growing» algorithm for matching of terrain images*”, *Image and Vision Computing*, 7(2):83–94, 1989.

⁶⁹ *Idem* 62.

⁷⁰ *Idem* 60.

⁷¹ Mikolajczyk, Krystian, Schmid, Cordelia, „*Scale & affine invariant interest point detectors*”, *International Journal of Computer Vision*, 60(1):63–86, 2004.

⁷² *Idem* 64.

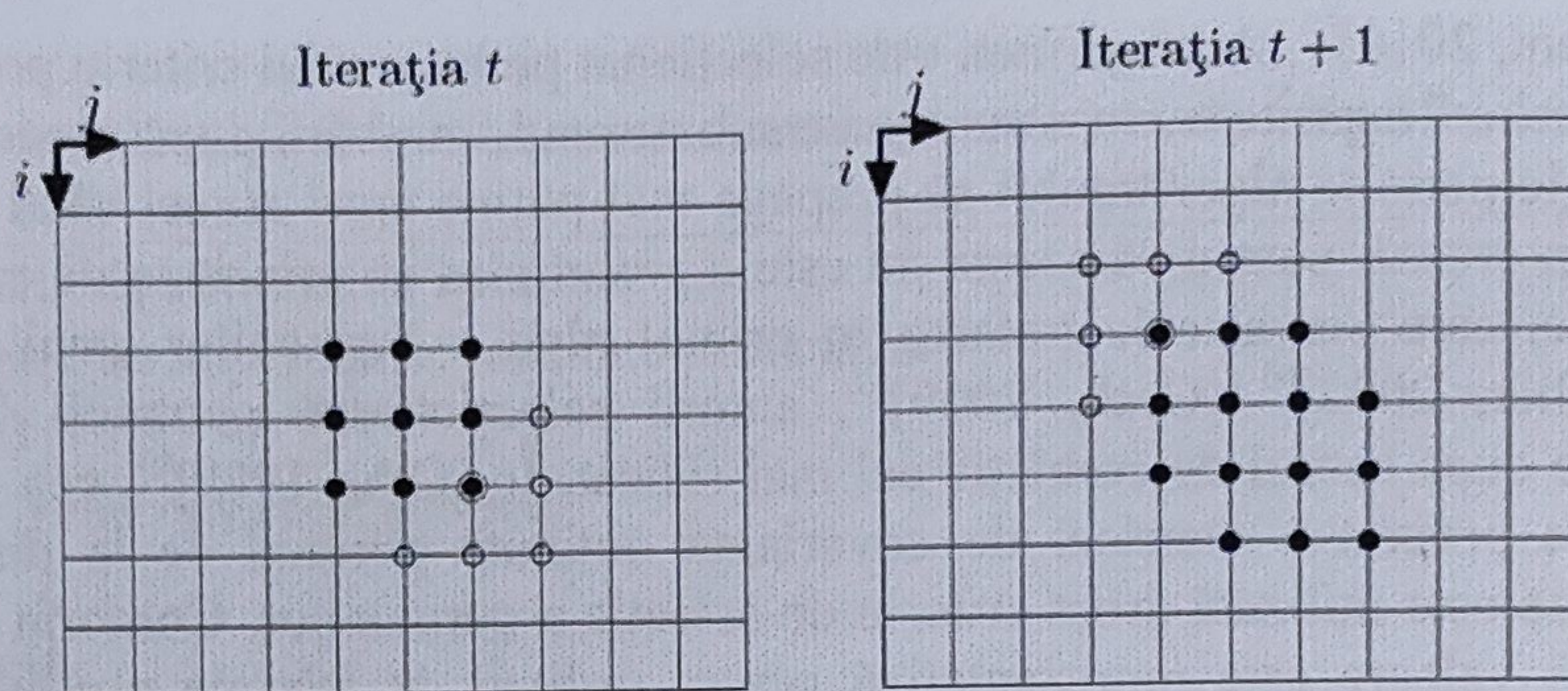


Fig. 21. *Abordarea secvențială a propagării* – un singur germen este selecționat în timpul unei aceleiași iterații (marcajele pixelilor sunt aceleași ca și în Fig. 20)

- 1: $\mathcal{G}_t \leftarrow$ setul inițial de germeni
- 2: $\mathcal{M} \leftarrow \mathcal{G}_t$
- 3: **Repetă**
- 4: selecționează germenul g din \mathcal{G} care satisface cel mai bine criteriul de selecție predefinit
- 5: $\mathcal{G} \leftarrow \mathcal{G} \setminus g$
- 6: **Execută pentru** toți pixelii neîmperecheați din vecinătatea germenului g considerat
- 7: caută în zona de căutare definită de germenul și pixelul curent candidatul care furnizează cel mai bun scor de corelație
- 8: **Dacă** acest candidat găsit formează o pereche care îndeplinește un ansamblu de condiții *a posteriori* **atunci**
- 9: Adaugă această pereche la \mathcal{M}
- 10: Adaugă această pereche la \mathcal{G}
- 11: **Încheiere dacă**
- 12: **Încheiere execută pentru**
- 13: **până când** $\mathcal{G} = \emptyset$
- 14: Returnează \mathcal{M}

ALGORITMUL 3 – Abordarea secvențială a propagării pentru punerea în corespondență a pixelilor

Algoritmul 3 prezintă abordarea secvențială a propagării pentru punerea în corespondență a pixelilor: la fiecare iterație, un singur germen este luat în considerare. Acesta este selecționat pe baza unui criteriu predefinit cu scopul de a propaga în primul rând germenii „cei mai buni”.

3.3 Condiții pentru propagare

Unul dintre inconvenientele metodelor cu propagare este riscul de a propaga erori pornind de la germeni incorecți. Este posibil să se limiteze acest efect utilizând condiții care previn propagarea atâta timp cât împerecherile realizate sunt considerate de mică încredere.

- **Condiția de omogenitate** – are ca scop de a preveni propagarea în zonele omogene sau prea puțin texturate pentru care lipsa informațiilor nu permite găsirea unor împerecheri într-o manieră credibilă cu o metodă locală [Lhuillier, 2002]⁷³. Propagarea unui germen este stopată atâta timp cât cei patru pixeli vecini ai acelui germen au valori apropiate, adică atunci când:

$$h(p_{i,j}^s) = \max \left\{ I^s \left(\begin{bmatrix} i+k & j+l \end{bmatrix}^T \right) - I^s \left(\begin{bmatrix} i & j \end{bmatrix}^T \right) \right\} < t_h \quad (32)$$

unde $t_h \in R$ este un prag și $(k,l) \in \{(1,0), (-1,0), (0,1), (0,-1)\}$.

- **Condiția de regiune** – presupune că, aproape peste tot, pixelii vecini care au valori apropiate sunt proiecții ale punctelor de pe aceeași suprafață și au disparități apropiate. Presupunem atunci că discontinuitățile de profunzime coincid cu contururile regiunilor formate din pixelii de valori omogene. Astfel, această condiție permite evitarea propagării unor disparități eronate în exteriorul regiunilor. În Fig. 22, căutarea corespondențelor vecinilor (cerculețe) pixelului stânga $g_{i,j}^s$ (cu negru) care aparține regiunii R_1 , se face în vecinătatea pixelului dreapta $g_{i,j}^d$ și zona de căutare (cu gri) se limitează la pixelii care aparțin aceleiași regiuni ca și pixelul dreapta al germenului. În acest exemplu, pixelii regiunii R_2' sunt excluși din zona de căutare a corespondentului pixelului pentru pixelul $p_{i+1,j-1}^s$. Totuși, utilizarea condiției de regiune implică prezența inițială a cel puțin unui germen în fiecare regiune pentru a putea găsi corespondenții pixelilor tuturor regiunilor. Această condiție este utilizată în [Zhuang, 2005]⁷⁴.

3.4 Îmbunătățirea propagării

Lucrările descrise în [Chambon, 2011]⁷⁵ au arătat că măsurile de corelație robuste la discontinuitățile de profunzime dau rezultate bune în acele zone, dar, de asemenea, că, în zonele fără dificultate, ele sunt mai puțin performante decât măsurile de corelație clasice. În consecință, cu scopul de a combina avantajele diferitelor

⁷³ Lhuillier, Maxime, Quan, Long, „Match propagation for image-based modeling and rendering”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(8): 1140–1146, 2002.

⁷⁴ Zhuang, Yong, Ma, Ma, Jun, Wei, Liu, Liu, „Progressive matching based on segmentation for 3D reconstruction”, în International Conference on Computer and Information Technology, pp. 575–579, Shanghai, China, septembrie 2005.

⁷⁵ Chambon, Sylvie, Crouzil, Alain, „Similarity measures for image matching despite occlusions in stereo vision”, Pattern Recognition, 44:2063–2075, 2011.

măsură de corelație, a fost propusă în [Gales, 2011]⁷⁶ o tehnică de punere în corespondență prin propagare multi-măsură. Ideea constă în inițierea prin propagarea germenilor, care *a priori* nu sunt situați în zonele de discontinuitate de profunzime, cu o măsură de corelație clasică și de a utiliza, asupra scorului de corelație, o condiție de prag strict pentru a opri propagare de îndată ce vecinătățile corespondențelor diferă puțin. Presupunem apoi că pixelii rămași de împerecheat sunt situați în zonele apropiate de discontinuitățile de profunzime și utilizăm o măsură de corelație robustă pentru a continua propagarea. Prin adaptarea măsurii de corelație la context, se obțin rezultate mai dense decât în cazul unei singure măsură fără să scadă procentajul împerecherilor corecte. În plus, tehnica multi-măsură se poate adapta la fel de bine abordării simultane ca și abordării secvențiale a punerii în corespondență prin propagare germenilor.

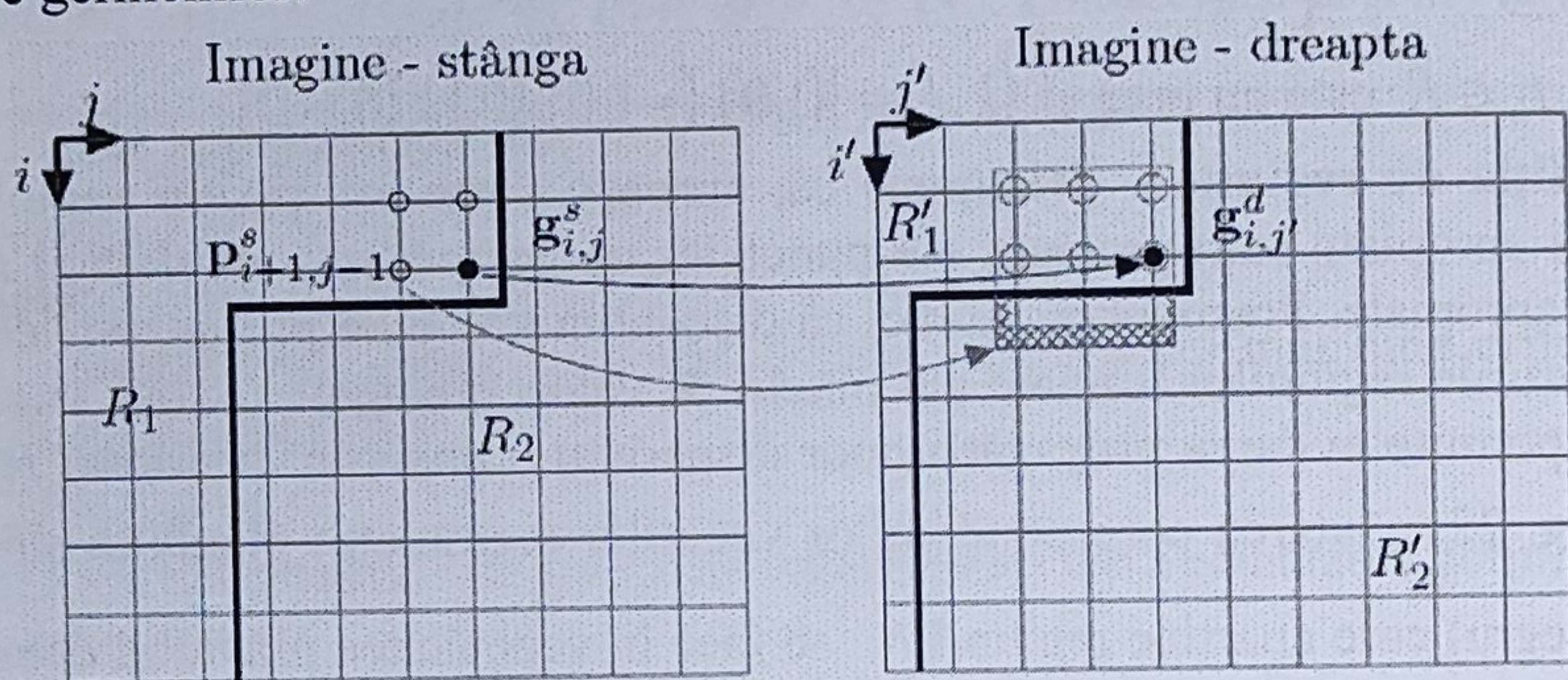


Fig. 22. Condiția de regiune

În cadrul abordării locale clasice, o idee similară a fost propusă în [Chambon, 2011(b)]⁷⁷. Sunt calculate independent două hărți de disparitate prin utilizarea a două măsuri de corelație diferite, o măsură clasică și o măsură robustă. Disparitățile finale sunt în obținute prin fuzionarea celor două hărți.

În abordarea secvențială (conform algoritmului 3), la fiecare iterație, „cel mai bun” germen trebuie să fie selecționat pentru a limita propagarea erorilor. În general, o măsură de corelație este utilizată ca și criteriu de selecție. O măsură de fiabilitate mai completă a fost propusă în [Gales, 2012]⁷⁸. Această măsură este calculată după propagarea unui germen și este constituită dintr-un termen de ambiguitate, care este calculat pornind de la măsura de corelație a gradientilor (conform ecuației 26), dintr-un termen de continuitate, care permite diminuarea influenței pixelilor candidați prea îndepărtați de germen, și dintr-un termen de consistență a culorilor, care favorizează

⁷⁶ Gales, Guillaume, „*Mise en correspondance de pixels pour la stéréovision binoculaire par propagation d'appariements de points d'intérêt et sondage de régions*”, teză de doctorat, Universitatea din Toulouse, iulie 2011.

⁷⁷ Chambon, Sylvie, Crouzil, Alain, „*Combination of correlation measures for dense stereo matching*”, în International Conference on Computer Vision Theory and Applications, material electronic, Algarve, Portugal, martie 2011(b).

⁷⁸ Gales, Guillaume, Chambon, Sylvie, Crouzil, Alain, McDonald, John, „*Reliability measure for propagation-based stereo matching*”, în Workshop on Image Analysis for Multimedia Interactive Services, Dublin City University, Irlanda, mai 2012.

pixelii a căror culoare este apropiată de cea a germenului. Fig. 23 prezintă harta de disparități obținute prin propagare în cazul perechii de imagini *Conuri* cu această măsură de fiabilitate.

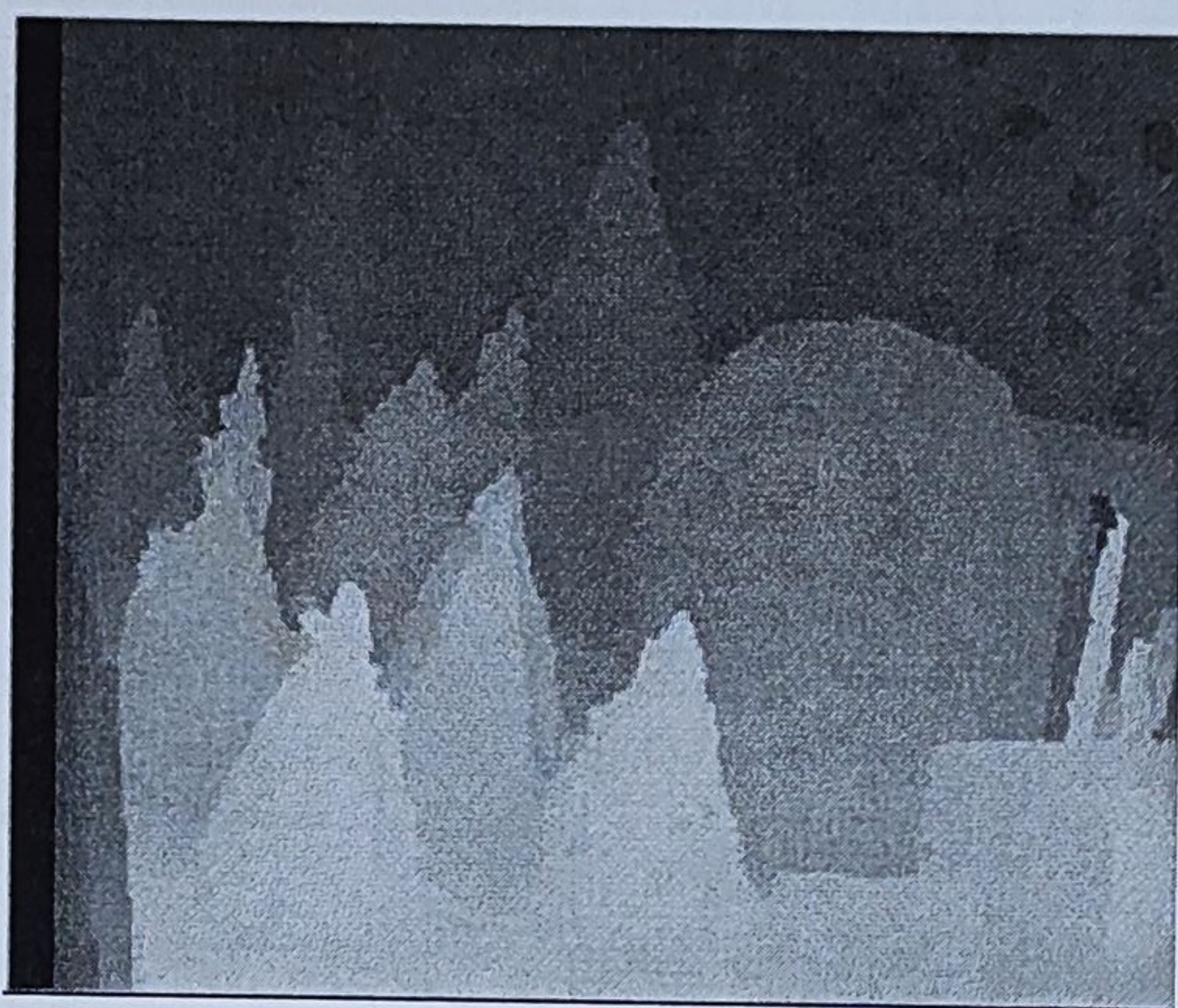


Fig. 23. *Harta disparităților perechii de imagini Conuri obținută prin propagarea cu o măsură de fiabilitate.*

4. Metode globale

În contradicție cu metodele locale care țin cont doar de vecinătatea fiecărui pixel, metodele globale încearcă să minimizeze un cost care evaluează eroarea globală de punere în corespondență, adică ținând cont de totalitatea imaginii. Se propune deci găsirea funcției de corespondență m care minimizează eroarea globală de punere în corespondență. Această eroare este cuantificată printr-o funcție de cost care integrează un termen de asociere între date, deci care face verificarea că valorile celor doi pixeli corespondenți sunt apropiate, și un termen de netezire, care permite asigurarea condiției limitei gradientului de disparitate. În Fig. 24, harta disparității este inițializată și apoi actualizată atât timp cât costul global al erorii de punere în corespondență, evaluat între necunoscute (harta de disparitate) și datele observate (imaginile stânga și dreapta), nu este considerat ca minim. În Fig. 25 se prezintă evoluția hărții de disparități a perechii de imagini *Conuri* în timpul procesului de punere în corespondență globală. Acest cost poate fi evaluat, deoarece, dacă harta de disparitate este corectă, atunci dacă se transformă imaginea stânga în funcție de aceste disparități, trebuie să obținem imaginea dreapta.

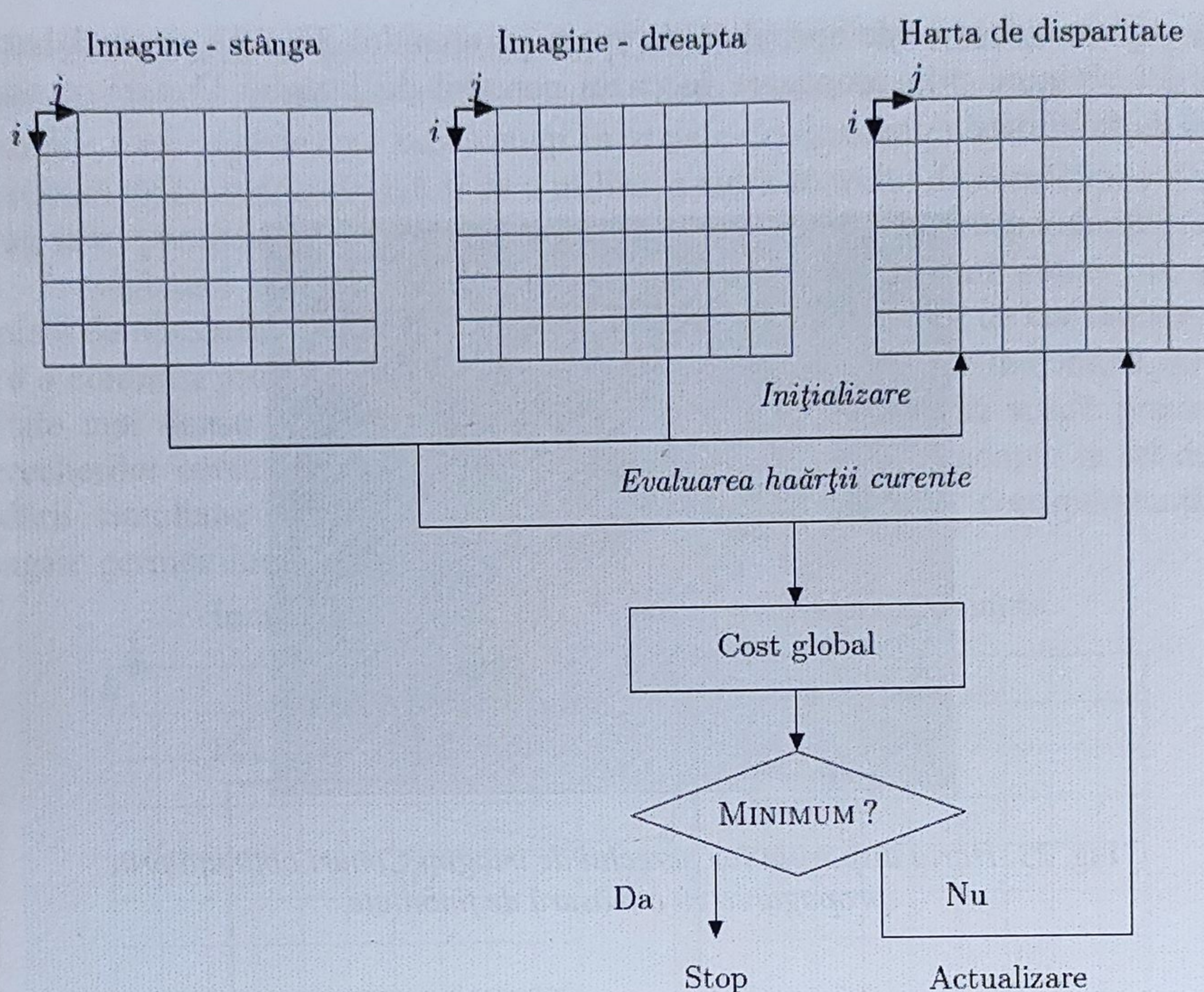


Fig. 24. Principiul metodelor globale

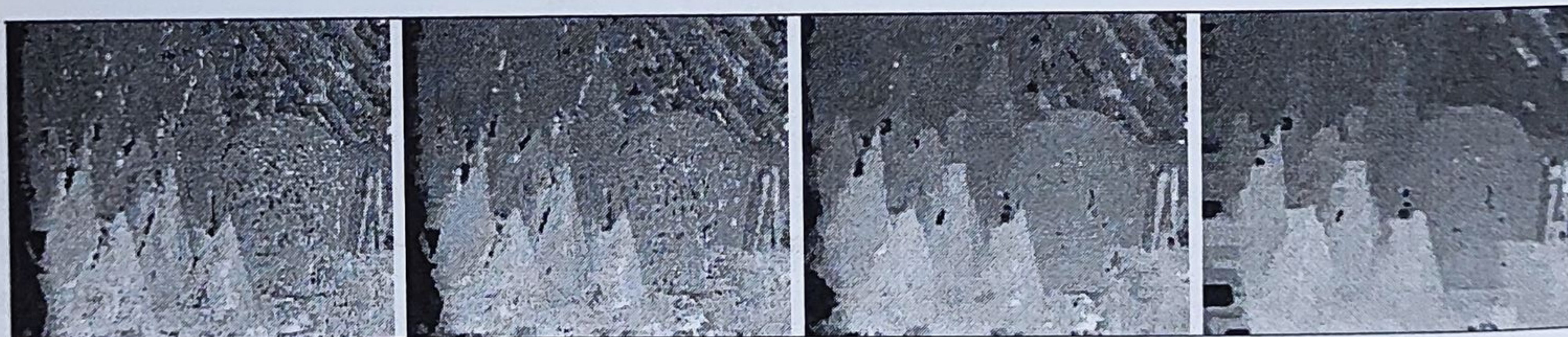


Fig. 25. Evoluția hărții de disparitate a perechii de imagini Conuri în timpul procesului de punere în corespondență globală (iterațiile 0, 2, 20 și 98)

4.1 Modelarea problemei cu câmpuri Markov

Problema punerii în corespondență stereoscopică poate fi modelată cu ajutorul câmpurilor Markov, ceea ce permite demonstrarea faptului că această problemă se poate rezolva prin minimizarea unei funcții de cost cu ajutorul metodelor de optimizare specifice [Geman, 1984]⁷⁹ [Szeliski, 1990]⁸⁰ [Li, 1994]⁸¹ [Freeman, 2000]⁸² [Veksler, 1999]⁸³ [Sun, 2002]⁸⁴.

⁷⁹ Geman, Stuart, Geman, Donald, „Stochastic relaxation, Gibbs distribution, and the bayesian restoration of images”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6(6):721–741, 1984.

Prezentarea problemei. Problema punerii în corespondență a pixelilor poate fi reconsiderată ca și o problemă de etichetare sau, în cazul imaginilor rectificate, pentru fiecare pixel, se dorește a fi atribuită o valoare de disparitate (eticheta). Astfel, dacă imaginea prelucrată este constituită din hw pixeli, și dacă numărul de valori posibile ale disparității este e , atunci există e^{hw} configurații posibile.

Notatii. În cele ce urmează se vor utiliza următoarele notații:

- Valorile pixelilor sunt reprezentate printr-un set de $n = hw$ variabile aleatoare $Y = \{Y_1, K, Y_n\}$.
- Valorile $Y_i, i \in [1, n]$ sunt notate cu y_i .
- Notăm cu γ o configurație a Y .
- Disparitățile sunt reprezentate printr-un set de variabile aleatoare $X = \{X_1, K, X_n\}$.
- Valorile X_i sunt notate cu x_i .
- Notăm cu χ o configurație a X .
- Notăm cu Ψ setul de configurații posibile pentru $X (\chi \in \Psi)$.

Valorile variabilelor Y_i fiind cunoscute, avem deci o configurație dată γ . Necunoscuta problemei este configurația cea mai probabilă χ a disparităților dacă se cunoaște γ ; o notăm cu $\hat{\chi}$. Deci soluția este configurația $\hat{\chi}$ pentru care probabilitatea $p(\chi|\gamma)$ este maximă:

$$\hat{\chi} = \arg \max_{\chi} p(\chi|\gamma) \quad (33)$$

Aceasta este ceea ce numim *maximum a posteriori* (MAP). Conform teoremei lui Bayes, ecuația (33) poate fi rescrisă astfel:

⁸⁰ Szeliski, Richard, „*Bayesian modeling of uncertainty in low-level vision*”, International Journal of Computer Vision, 5(3):271–302, decembrie 1990.

⁸¹ Li, Stan Z., „*Markov random field models in computer vision*”, în European Conference on Computer Vision, vol. 2, pp. 361–370, Stockholm, Suedia, mai 1994.

⁸² Freeman, William. T., Pasztor, Egon C., Carmichael, Owen T., „*Learning low-level vision*”, International Journal of Computer Vision, 40(1):25–47, 2000.

⁸³ Veksler, Olga, „*Efficient Graph-Based Energy Minimization Methods in Computer Vision*”, teză de doctorat, Cornell University, Ithaca, NY, S.U.A., august 1999.

⁸⁴ Sun, Jian, Shum, Heung-Yeung, Zheng, Nan-Ning, „*Stereo matching using belief propagation*”, în European Conference on Computer Vision, pp. 450–452, Copenhaga, Danemarca, mai 2002.

$$\hat{\chi} = \arg \max_{\chi} p(\gamma|\chi)p(\chi) \quad (34)$$

Prin definiție, graful care reprezintă variabilele aleatoare ale X constituie un câmp Markov pentru că el îndeplinește proprietățile următoare:

$$\begin{cases} p(\chi) > 0 & \forall \chi \in \Psi \\ p(x_i|x_j, j \neq i) = p(x_i|x_v) & X_v \in N_{X_i}^4 \end{cases} \quad (35)$$

unde $N_{X_i}^4$ este setul de variabile aleatoare 4-vecini a lui X_i în graf (conform Fig. 26). Altfel spus, starea unei variabile aleatoare nu poate depinde decât de starea vecinilor săi direcți.

Pentru un câmp Markov, este apoi posibil să se calculeze probabilitatea unei configurații χ grație teoremei Hammersley-Clifford care arată echivalența unui câmp Markov cu câmpul lui Gibbs, de unde probabilitatea unei configurații χ :

$$p(\chi) = \frac{1}{z} \exp \left(- \sum_{\{X_p, X_q\} \in N} V(x_p, x_q) \right) \quad (36)$$

unde N conține mulțimea tuturor cuplurilor vecine două câte două în graf, unde z este un termen de normalizare și unde V este o funcție care cuantifică variația valorilor x_p și x_q :

$$\begin{aligned} V: R^2 &\rightarrow R \\ (x_p, x_q) &\propto V(x_p, x_q) \end{aligned} \quad (37)$$

Având determinată $p(\chi)$, pentru a putea calcula $\hat{\chi}$ conform ecuației (34), rămâne de determinat $p(\gamma|\chi)$. Este un termen care cuantifică probabilitatea de a obține valoarea unui pixel dacă se cunoaște disparitatea sa. Acesta nu se poate calcula direct. În schimb, dacă valoarea disparității este corectă, ar trebui să găsim în cealaltă imagine un pixel de valoare apropiată. Calculând diferența dintre valorile celor doi pixeli, se obține un scor invers proporțional cu ceea ce dorim să cuantizăm. Se poate deci scrie:

$$p(\gamma|\chi) \propto \exp \left(- \sum_{X_p} D(x_p) \right). \quad (38)$$

unde \propto semnifică faptul că cei doi termeni sunt proporționali și D este o funcție care, pornind de la o disparitate estimată, cuantifică similaritatea dintre cei doi pixeli presupuși corespondenți:

$$\begin{aligned} D: R &\rightarrow R \\ x_p &\propto D(x_p) \end{aligned} \quad (39)$$

În consecință, pornind de la ecuațiile (34), (36) și (38), avem:

$$\hat{\chi} = \arg \max_{\chi} \exp \left(- \sum_{X_p \in X} D(x_p) - \sum_{\{X_p, X_q\} \in N} V(x_p, x_q) \right) \quad (40)$$

ceea ce ne readuce la a căuta configurația χ care minimizează funcția de cost de forma:

$$E(\chi) = \sum_{X_p \in X} D(x_p) + \sum_{\substack{\{X_p, X_q\} \in N \\ \chi}} V(x_p, x_q). \quad (41)$$

Modelul construit poate fi reprezentat printr-un graf (conform Fig. 26). Cerculețele reprezintă variabilele aleatoare. Fiecărui cerculeț negru care corespunde unei variabile aleatoare $Y_i \in Y$ (ea reprezintă o valoare dată observată) i se asociază o variabilă aleatoare $X_i \in X$ (aceasta reprezintă necunoscuta). În plus, există legături între $X_i \in X$ vecini care impun ca valorile variabilelor aleatoare a doi vecini să nu fie prea diferite. Variabilele aleatoare X constituie un câmp Markov pentru că ele sunt legate, dar fiecare nu poate depinde decât de vecinii săi direcți, cum este cazul, de exemplu, variabilei aleatoare marcată cu gri închis a cărei stări nu poate depinde decât de cele marcate cu gri deschis.

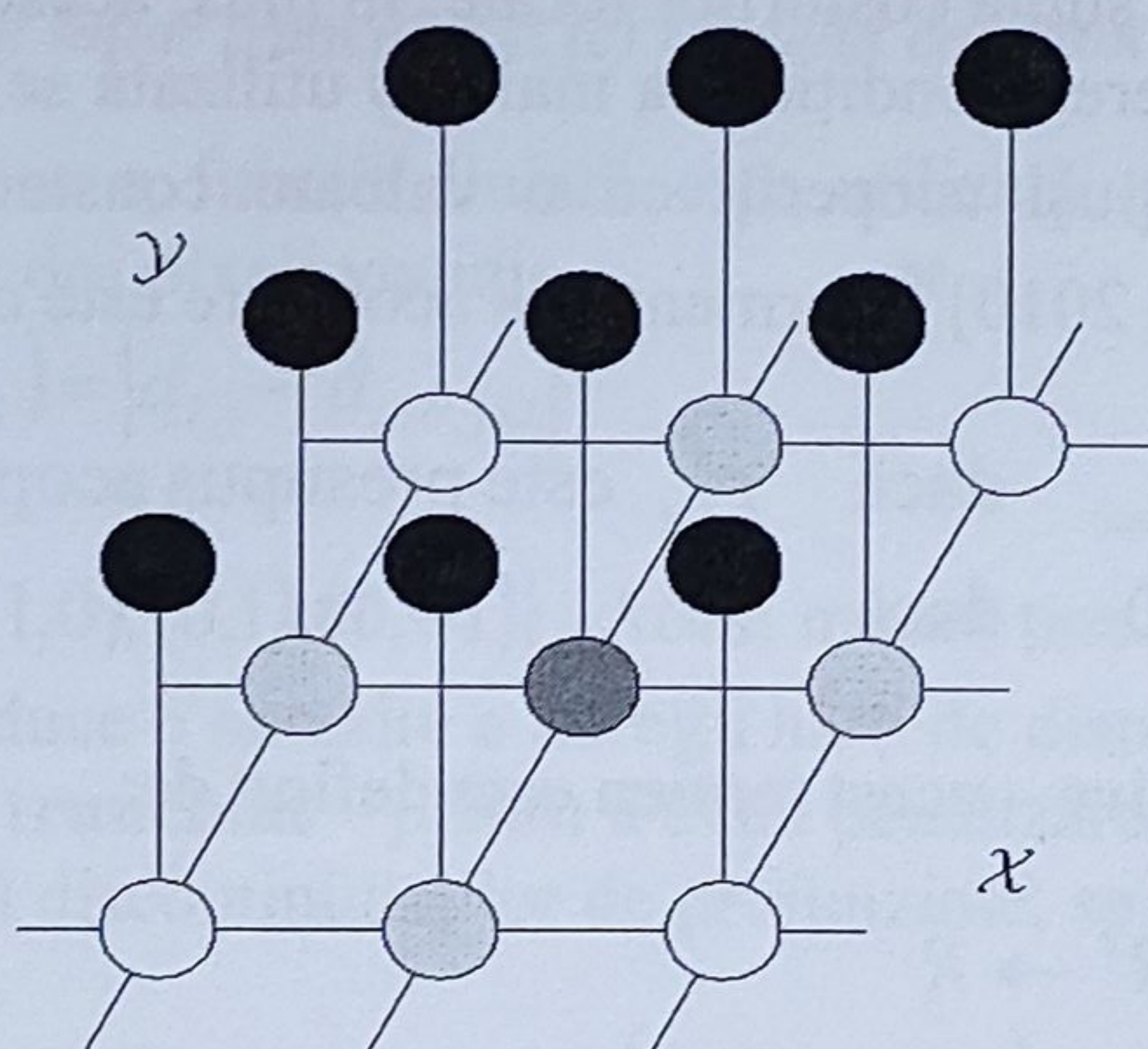


Fig. 26. Modelarea problemei cu ajutorul câmpurilor Markov

4.2 Funcții de cost

Funcția de cost care exprimă eroarea globală de punere în corespondență trebuie să aibă aceeași formă ca și funcția dată de ecuația (41) în care fiecare variabilă aleatoare din X corespunde disparității asociate unui pixel. În felul acesta putem considera că o configurație χ poate fi reprezentată printr-o funcție de disparitate d . Pe de altă parte, datele observate, adică valorile variabilelor aleatoare din Y , corespund valorilor pixelilor din cele două imagini. În consecință, se poate rescrie funcția de cost (41) în maniera următoare:

$$\underbrace{E(\chi)}_{C(d^s \rightarrow d)} = \underbrace{\sum_{x_p \in X} D(x_p)}_{\sum_{i=0}^{h-1} \sum_{j=0}^{w-1} C_{atasament}(d_{i,j}^s)} + \underbrace{\sum_{\{x_p, x_q\} \in N} V(x_p, x_q)}_{\lambda \sum_{i=0}^{h-1} \sum_{j=0}^{w-1} \sum_{k,l} C_n(d_{i,j}^s, d_{i+k,j+l}^s)} \quad (42)$$

cu $(k, l) \in \{(1, 0), (-1, 0), (0, 1), (0, -1)\}$ și unde parametrul $\lambda \in R$ permite reglarea influenței termenului de netezire față de termenul de asociere la date.

- **Termenul de asociere de date** – este definit prin:

$$\begin{aligned} C_{atasament} &: R \rightarrow R \\ d_{i,j}^s &\propto C_{atasament}(d_{i,j}^s) \end{aligned} \quad (43)$$

Acesta evaluează potrivirea între disparitățile date de funcția de disparitate și datele observate.

El corespunde în general unei măsuri de corelație între valorile pixelilor din imaginea de referință și valorile pixelilor corespondenți date de disparitățile estimate. Unele metode utilizează suma costurilor locale. În plus, acest termen poate să includă un termen de constrângere. Condiția cea mai des utilizată se referă la acoperiri. Sunt penalizați pixelii presupuși acoperiți cu o valoare constantă predefinită λ_{occ} . În [Wang, 2008]⁸⁵ [Bleyer, 2010]⁸⁶, termenul de acoperire este dat de:

$$C_{occ}(d_{i,j}^s) = \begin{cases} \lambda_{occ} & \text{dacă } p_{i,j}^s \text{ este presupus acoperit} \\ 0 & \text{dacă } \quad \quad \quad \text{nu.} \end{cases} \quad (44)$$

- **Termenul de netezire** – acest termen este definit de:

$$\begin{aligned} C_n &: R^2 \rightarrow R \\ d_{i,j}^s, d_{i+k,j+l}^s &\propto C_n(d_{i,j}^s, d_{i+k,j+l}^s) \end{aligned} \quad (45)$$

Acesta asigură disparități apropiate pentru pixelii vecini, ceea ce este cazul aproape peste tot, mai puțin la nivelul discontinuităților de profunzime. Modelele funcției de netezire care sunt cel mai mult utilizate în problemele de punere în corespondență a pixelilor sunt în special cele prezentate în [Veksler, 1999]⁸⁷:

⁸⁵ Wang, Zeng-Fu, Zheng, Zhi-Gang, „A region based stereo matching algorithm using cooperative optimization”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8, Anchorage, AK, S.U.A., iunie 2008.

⁸⁶ Bleyer, Michael, Rother, Carsten, Kohli, Pushmeet, „Surface stereo with soft segmentation”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1570–1577, San Francisco, CA, S.U.A., iunie 2010.

⁸⁷ Veksler, Olga, „Efficient Graph-Based Energy Minimization Methods in Computer Vision”, teză de doctorat, Cornell University, Ithaca, NY, S.U.A., august 1999.

Stabilirea corespondențelor între pixeli în stereoviziunea binoculară

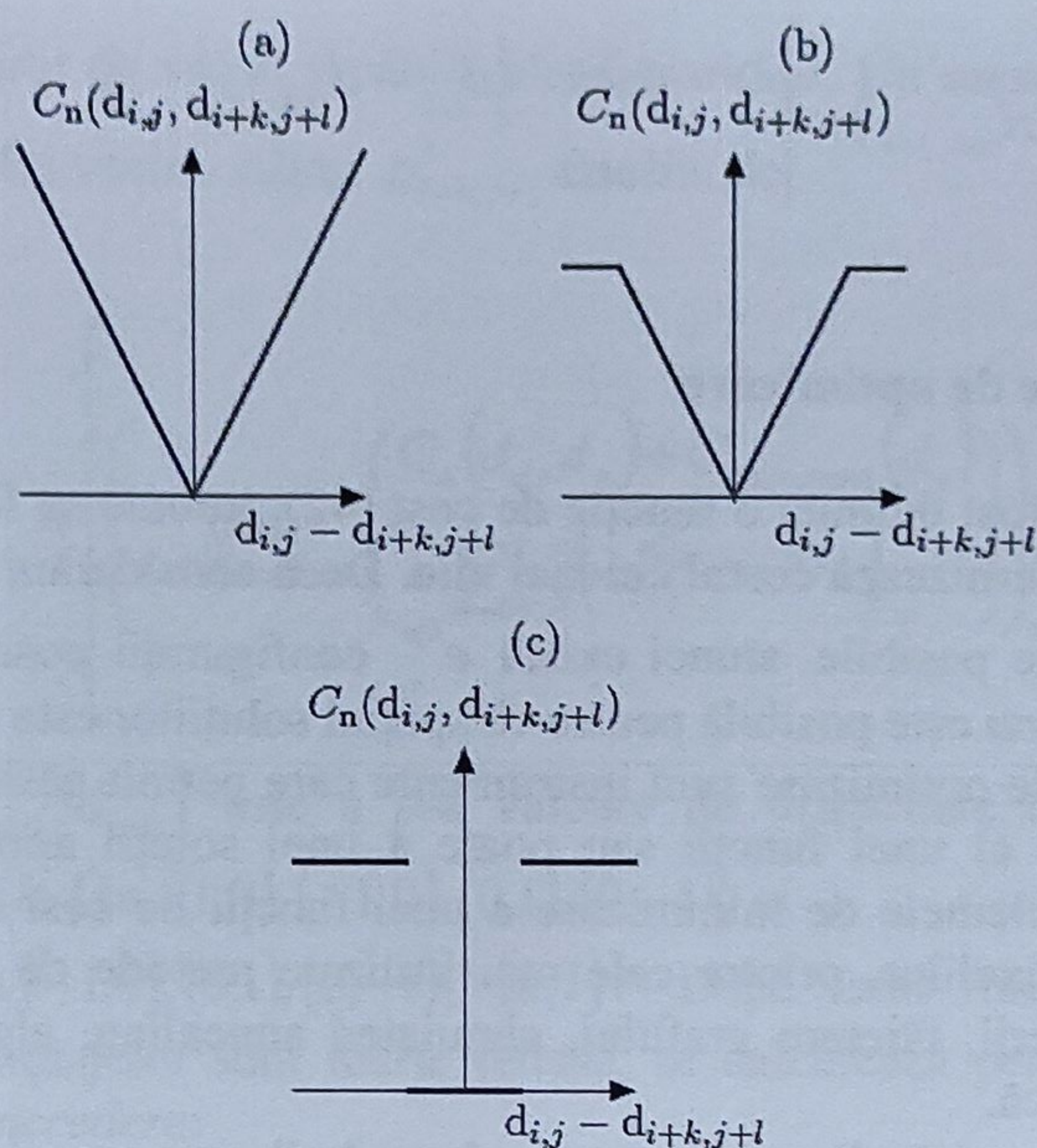


Fig. 27. Modele de funcții de netezire – (a) modelul liniar, (b) modelul liniar trunchiat și (c) modelul constant pe porțiuni

- **Modelul liniar** – o funcție liniară este utilizată pentru a măsura variația disparității între doi pixeli vecini:

$$C_n(d_{i,j}, d_{i+k,j+l}) = |d_{i,j} - d_{i+k,j+l}| \quad (46)$$

cu $(k,l) \in \{(1,0), (-1,0), (0,1), (0,-1)\}$. Acest model ține cont de discontinuitățile de profunzime și induce o netezire a întregii hărți de disparitate.

- **Modelul liniar trunchiat** – pentru a evita penalizarea exagerată a costului global la nivelul discontinuităților de profunzime, se poate utiliza modelul liniar trunchiat:

$$C_n(d_{i,j}, d_{i+k,j+l}) = \begin{cases} |d_{i,j} - d_{i+k,j+l}| & \text{dacă } |d_{i,j} - d_{i+k,j+l}| < \lambda_l \\ \lambda_l & \text{dacă nu.} \end{cases} \quad (47)$$

unde $\lambda_l \in R$ este un prag (conform Fig. 27).

- **Modelul quadratic trunchiat** – este o alternativă la modelul liniar:

$$C_n(d_{i,j}, d_{i+k,j+l}) = \begin{cases} (d_{i,j} - d_{i+k,j+l})^2 & \text{dacă } (d_{i,j} - d_{i+k,j+l})^2 < \lambda_l \\ \lambda_l & \text{dacă nu.} \end{cases} \quad (48)$$

- **Modelul constant pe porțiuni** – acest model nu impune nicio penalitate pentru variațiile mici ale disparității între doi vecini:

$$C_n(d_{i,j}, d_{i+k,j+l}) = \begin{cases} 0 & \text{dacă } |d_{i,j} - d_{i+k,j+l}| < \lambda_l \\ \lambda_l & \text{dacă nu.} \end{cases} \quad (49)$$

4.3 Metode de optimizare

După ce a fost definită o funcție de cost (42), trebuie să fie găsită funcția de disparitate d , care furnizează costul cel mai mic. Dacă considerăm un număr fix e de valori de disparitate posibile, atunci există e^{hw} configurații posibile. În practică, o căutare exhaustivă nu este posibilă pentru că spațiul soluțiilor este mult prea mare.

Metodele de optimizare sunt instrumente care permit găsirea minimului (sau maximului) global al unei funcții sau poate a unei soluții apropiate de minimul căutat. Pentru problemele de minimizare a unei funcții de cost în cazul punerii în corespondență a pixelilor, printre cele mai utilizate metode de optimizare se află propagarea încrederii, tăietura grafului, simulated annealing, algoritmi genetici și propagarea dinamică.

- **Propagarea încrederii** – propagarea încrederii este un algoritm introdus în [Pearl, 1982]⁸⁸ care permite deducerea, pornind de la un câmp Markov, a legii marginale în fiecare nod al grafului. Utilizarea sa pentru problema de punere în corespondență a pixelilor a fost introdusă în [Sun, 2002]⁸⁹ și este de asemenea utilizată în [Klaus, 2006]⁹⁰ [Yang, 2009]⁹¹ [Xu, 2008]⁹² [Yang, 2008]⁹³ [Sun, 2005]⁹⁴ [Zitnick, 2007]⁹⁵ [Taguchi, 2008]⁹⁶. Această metodă constă în determinarea pentru fiecare pixel a densității disparității și în selecționarea valorii celei mai probabile. Pentru aceasta se utilizează un algoritm iterativ. La fiecare iterație, pixelii vecini schimbă mesaje, sub forma unor vectori de dimensiuni

⁸⁸ Pearl, Judea, „*Reverend Bayes on inference engines: A distributed hierarchical approach*”, în National Conference on Artificial Intelligence, pag. 133–136, Pittsburgh, PA, S.U.A., august 1982.

⁸⁹ Sun, Jian, Shum, Heung-Yeung, Zheng, Nan-Ning, „*Stereo matching using belief propagation*”, în European Conference on Computer Vision, pp. 450–452, Copenhaga, Danemarca, mai 2002.

⁹⁰ Klaus, Andreas, Sormann, Mario, Karner, Konrad, „*Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure*”, în International Conference on Pattern Recognition, vol. 3, pp. 15–18, Hong Kong, august 2006.

⁹¹ Yang, Qingxiong, Wang, Ling, Yang, Ruigang, Stewenius, Henrik, Nistér, David, „*Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(3):492–504, 2009.

⁹² Xu, Li, Jia, Jiaya, „*Stereo matching: an outlier confidence approach*”, în European Conference on Computer Vision, vol. 4, pp. 775–787, Marsilia, Franța, octombrie 2008.

⁹³ Yang, Qingxiong, Engels, Chris, Akbarzadeh, Amir, „*Near real-time stereo for weakly-textured scenes*”, în British Machine Vision Conference, pp. 80–87, Leeds, U.K., septembrie 2008.

⁹⁴ Sun, Jian, Li, Yin, Kang, Sing B., Shum, Heung-Yeung, „*Symmetric stereo matching for occlusion handling*”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 399–406, San Diego, CA, S.U.A., iunie 2005.

⁹⁵ Zitnick, C. Lawrence, Kang, Sing B., „*Stereo for image-based rendering using oversegmentation*”, International Journal of Computer Vision, 75(1):49–65, 2007.

⁹⁶ Taguchi, Yuichi, Wiburn, Bennett, Zitnick, C. Lawrence, „*Stereo reconstruction with mixed pixels using adaptive over-segmentation*”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8, Anchorage, AK, S.U.A., iunie 2008.

egale cu numărul de valori posibile ale disparității. Un mesaj $m_t^{p_{i,j}^s \rightarrow p_{i+k,j+l}^s}$ de la iterația t între doi vecini $p_{i,j}^s$ și $p_{i+k,j+l}^s$ este dat de:

$$m_t^{p_{i,j}^s \rightarrow p_{i+k,j+l}^s} = \left[\begin{array}{c} \min_{d_q \in [d_{\min}, d_{\max}]} \left(C_n(d_q, d_p) + C_{\text{atasament}}(d_q) \right) \\ + \sum_{k',l'} \left(m_{t-1}^{p_{i+k',j+l'}^s \rightarrow p_{i,j}^s} \right)_{d_q} \\ \vdots \end{array} \right]_{d_p} \quad (50)$$

unde $d_p \in [d_{\min}, d_{\max}]$ este a p -a valoare de disparitate din setul de valori posibile și cu $(k,l) \in \{(1,0), (-1,0), (0,1), (0,-1)\}$ și $(k',l') \in \{(1,0), (-1,0), (0,1), (0,-1)\} \setminus \{k,l\}$. După mai multe iterații, când modificările mesajelor sunt foarte reduse, se calculează pentru fiecare nod un vector numit *încredere*:

$$b^{p_{i,j}^s} = \left[\begin{array}{c} C_{\text{atasament}}(d_p) + \sum_{k,l} \left(m_t^{p_{i,j}^s \rightarrow p_{i+k,j+l}^s} \right)_{d_p} \\ M \end{array} \right]. \quad (51)$$

În sfârșit, valoarea de disparitate finală este dată de:

$$d_{i,j} \leftarrow \arg \min_{d_p} b^{p_{i,j}^s}. \quad (52)$$

- **Tăietura grafului** – În primul rând, problema pe care o dorim a fi rezolvată este reprezentată printr-un graf orientat. Există două noduri particulare: de intrare și de ieșire. Este necesară determinarea unei tăieturi minimale a acestui graf, adică un set de arcuri astfel ca, fără aceste arcuri, să nu mai existe în cadrul grafului o cale între ieșire și sursă și astfel ca suma ponderilor arcelor tăiate să fie cât mai mică, ceea ce implică un flux maximal între sursă și ieșire conform teoremei lui Ford și Fulkerson [Ford, 1957]⁹⁷. Algoritmul lui Ford și Fulkerson permite determinarea fluxului maximal (conform algoritmului 4). Algoritmul 5, introdus în [Veksler, 1999]⁹⁸, este utilizat în special pentru calculul disparităților, pornind de la modele diferite de grafuri. Este un algoritm iterativ care se aplică unui *move space* care definește schimbările de configurații. De exemplu, „ $\alpha - \beta$ -relabel move space” constă, în configurația curentă, din înlocuirea valorilor tuturor variabilelor de valoare α cu valoarea β . În [Veksler, 1999]⁹⁹ sunt propuse diferite

⁹⁷ Ford, Lester R. Jr., Fulkerson, Delbert R., „A simple algorithm for finding maximal network flows and an application to the hitchcock problem”, Canadian Journal of Mathematics, 9:210–218, 1957.

⁹⁸ Veksler, Olga, „Efficient Graph-Based Energy Minimization Methods in Computer Vision”, teză de doctorat, Cornell University, Ithaca, NY, S.U.A., august 1999.

⁹⁹ Idem.

modele de *move space* și de grafuri pentru probleme de computer vision, printre care și punerea în corespondență a pixelilor. Aceste modele diferă în special prin modelul de funcție utilizată în cadrul termenului de netezire. În [Roy, 1999]¹⁰⁰, problema punerii în corespondență a pixelilor este formalizată pentru cazul unei achiziții multi-view (mai mult de 2 camere de luat vederi) și fără utilizarea condiției de epipolaritate.

- 1: **Repetă**
- 2: selecționează la întâmplare un drum oarecare de la sursă la destinație (dacă ponderea unui arc este nulă, drumul nu poate fi ales).
- 3: determină fluxul maxim f_m de-a lungul parcursului selectat. Acesta corespunde ponderii arcului cel mai slab.
- 4: extrage f_m prin scădere din ponderile tuturor arcelor de pe parcursul considerat.
- 5: **până când** nu mai este nici un parcurs posibil
- 6: Caută tăietura minimă printre toate arcele a căror pondere este nulă.

ALGORITMUL 4. – Algoritmul lui Ford și Fulkerson – acesta permite găsirea arcelor care dau fluxul maximal și tăietura minimală a unui graf.

- 1: $d_0 \leftarrow$ stare inițială oarecare
- 2: C : funcție de cost
- 3: $\hat{d} \leftarrow d_0$, \hat{d} conține rezultatul
- 4: **Atât timp cât** „adevărat” **execută**
- 5: găsește configurația „vecină” minimală $d_i \leftarrow \arg \min_{d'} C(d')$.
 Configurațiile „vecine” sunt date de o funcție *move space*. Această configurație minimală este calculată prin tăietura grafului.
- 6: **Dacă** $C(d_i) < C(\hat{d})$ **atunci**
- 7: $\hat{d} \leftarrow d_i$
- 8: **Dacă nu**
- 9: **returnează** \hat{d}
- 10: **sfârșit dacă**
- 11: **sfârșit atâta timp cât**

ALGORITMUL 5. – Algoritmul Veksler.

¹⁰⁰ Roy, Sébastien, „Stereo without epipolar lines: A maximum-flow formulation”, International Journal of Computer Vision, 34(2-3):147-161, 1999.

- **Călirea simulată** – „Simulated annealing” este un algoritm iterativ și stocastic, adică bazat pe extrageri aleatoare (conform algoritmului 6). Este în special utilizată în [Park, 2007]¹⁰¹. La fiecare iterație, o nouă configurație de disparități este aleasă aleator. Dacă costul acestei noi configurații este mai scăzut decât cel mai bun cost curent, atunci ea este acceptată. În caz contrar, ea este acceptată în funcție de o anumită probabilitate care depinde de o variabilă numită *temperatură*, și aceasta cu scopul de a evita ca algoritmul să nu convergă către un minim local. Pe de altă parte, la fiecare iterație, o funcție face să scadă *temperatura* pentru ca probabilitatea de acceptare să devină din ce în ce mai scăzută.

```
1:  $d_0 \leftarrow$  stare inițială oarecare
2:  $T_0 \leftarrow$  temperatura inițială
3:  $T \leftarrow T_0$ 
4:  $f_u$ : funcția de disparitate curentă
5:  $f_d$ : funcția de descreștere a temperaturii
6: Repetă
7:    $d_t \leftarrow f_u(d_{t-1})$ 
8:    $T \leftarrow f_d(T)$ 
9:    $\Delta C = C(d_t) - C(d_{t-1})$ , variația costului
10:  Dacă  $\Delta C < 0$  atunci
11:    acceptă schimbarea
12:  dacă nu
13:    acceptă schimbarea cu o probabilitate ce este în funcție de
    temperatură
14:  sfârșit dacă
15: până la convergență
16: returnează  $d_t$ 
```

ALGORITMUL 6. – Algoritmul „simulated annealing” pentru minimizarea unei funcții de cost a funcției de disparitate. Funcția f_u menține actualizată funcția de disparitate curentă.

- **Algoritmi genetici** – Această tehnică este introdusă pentru cazul general în [Holland, 1975]¹⁰². Problema de rezolvat este modelată printr-o populație care regroupează o mulțime de indivizi. Fiecare individ posedă o hartă genetică, adică un set de proprietăți. Indivizii cei mai bine adaptați sunt apoi selecționați cu

¹⁰¹ Park, Joonyoung, Kim, Wonsik, Lee, Kyoung M., „Stereo matching using population based MCMC”, în Asian Conference on Computer Vision, vol. 2, pp. 560–569, Tokyo, Japonia, noiembrie 2007.

¹⁰² Holland, John H., „Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence”, MIT Press, 1975.

ajutorul unei funcții de *fitness* și se reproduc prin amestecarea genelor lor. Ideea este de a obține noi indivizi cu cele mai bune proprietăți ale celor doi părinți. De asemenea pot interveni și mutații aleatoare. Astfel, se creează o nouă populație cu un nivel de *fitness* mai ridicat. Procesul este repetat până când se ajunge la apariția unei populații optimale, adică o populație care nu mai evoluează sau aproape că nu mai evoluează. În cazul punerii în corespondență a pixelilor, un individ corespunde unei funcții de disparitate ale cărei proprietăți sunt diferitele valori de disparitate asociate fiecărui pixel. Selecția celor mai bune funcții de disparitate este bazată pe funcția de cost global. Procesul este repetat de mai multe ori și din aceasta rezultă populația „optimală”. În sfârșit, cel mai bun individ al populației apărute, adică cea mai bună funcție de disparitate, este selecționată ca rezultat.

- 1: $D_0 \leftarrow \{K, d_k, K\}$, stare inițială oarecare (generarea unei populații inițiale de $k = 0, K, n$ indivizi)
- 2: **Repetă**
- 3: selecționează cei mai buni m indivizi din populația D conform funcție de *fitness*
- 4: încrucișează cei mai buni indivizi între ei pentru a genera o nouă populație. Partiționarea genelor se face la întâmplare
- 5: efectuează mutații aleatoare
- 5: **până când** apare o populație care satisface cel mai bine criteriul de *fitness*
- 6: **returnează** cel mai bun individ d al ultimei populații

ALGORITMUL 7. – Algoritmul genetic – acest algoritm efectuează încrucișări între indivizi și selecționează copii care satisfac cel mai bine funcția de cost (*fitness*).

- **Programare dinamică** – Această tehnică constă în calcularea pentru fiecare linie, a drumului care dă costul minim dintr-o matrice care listează toate costurile între toate împerecherile posibile ale liniei considerate. Totuși, această tehnică nu ține cont de relațiile liniilor unele cu altele. Metoda construiește pentru fiecare linie a imaginii de referință o matrice care conține toate costurile de punere în corespondență. Apoi este nevoie de găsirea unui drum în această matrice care minimizează funcția de cost.

Informații complementare se găsesc în [Scharstein, 2002]¹⁰³ [Szeliski, 2010]¹⁰⁴.

¹⁰³ Scharstein, Daniel, Szeliski, Richard, „A taxonomy and evaluation of dense twoframe stereo correspondence algorithms”, International Journal of Computer Vision, 47(1):7–42, 2002.

¹⁰⁴ Szeliski, Richard, „Computer Vision: Algorithms and Applications”, Springer, 2010.

5. Metode bazate pe regiuni

Conform protocolului de evaluare a lui Middlebury¹⁰⁵, care compară metode de punere în corespondență densă pentru stereoscopia binoculară, cele mai multe metode bine clasificate sunt bazate pe regiuni.

Principiul metodelor care utilizează regiunile se bazează pe ipoteza că, într-o imagine, o regiune de culoare omogenă suficient de mică poate fi proiecția unei aceleiași suprafețe a scenei și că acea suprafață poate fi abordată printr-un model în spațiul disparităților (de exemplu un plan sau o suprafață *B-Spline*) [Hong, 2004]¹⁰⁶ [Sun, 2005]¹⁰⁷ [Klaus, 2006]¹⁰⁸ [Wang, 2008]¹⁰⁹ [Yang, 2008]¹¹⁰ [Yang, 2009]¹¹¹ [Taguchi, 2008]¹¹² [Bleyer, 2010]¹¹³. Utilizarea acestor regiuni poate ajuta punerea în corespondență a regiunilor omogene și permite, de asemenea, evitarea artefactelor la nivelul discontinuităților de profunzime. Modelul plan este relativ simplu și asigură o bună aproximare, în particular pentru cele mai mici regiuni, dar poate deveni imprecis atunci când regiunile nu sunt proiecțiile unor fațete plane. În acest ultim caz, modelul *B-Spline* poate da o aproximare mai bună. Cu toate acestea, el este mai complicat de configurat (ordinul și vectorul nodal trebuie alese cu grijă) și pot apărea oscilații importante ale rezultatului [Lin, 2003]¹¹⁴ [Bleyer, 2010]¹¹⁵.

Metodele bazate pe regiuni implică parcurgerea mai multor etape prezentate pe scurt în Fig. 28 și pe care le vom descrie în continuarea acestui capitol.

¹⁰⁵ <http://vision.middlebury.edu/stereo>

¹⁰⁶ Hong, Li, Chen, George, „*Segment-based stereo matching using graph cuts*”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 74–81, Washington, DC, S.U.A., iunie 2004.

¹⁰⁷ Sun, Jian, Li, Yin, Kang, Sing B., Shum, Heung-Yeung, „*Symmetric stereo matching for occlusion handling*”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 399–406, San Diego, CA, S.U.A., iunie 2005.

¹⁰⁸ Klaus, Andreas, Sormann, Mario, Karner, Konrad, „*Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure*”, în International Conference on Pattern Recognition, vol. 3, pp. 15–18, Hong Kong, august 2006.

¹⁰⁹ Wang, Zeng-Fu, Zheng, Zhi-Gang, „*A region based stereo matching algorithm using cooperative optimization*”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8, Anchorage, AK, S.U.A., iunie 2008.

¹¹⁰ Yang, Qingxiong, Engels, Chris, Akbarzadeh, Amir, „*Near real-time stereo for weakly-textured scenes*”, în British Machine Vision Conference, pp. 80–87, Leeds, U.K., septembrie 2008.

¹¹¹ Yang, Qingxiong, Wang, Ling, Yang, Ruigang, Stewenius, Henrik, Nistér, David, „*Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(3):492–504, 2009.

¹¹² Taguchi, Yuichi, Wiburn, Bennett, Zitnick, C. Lawrence, „*Stereo reconstruction with mixed pixels using adaptive over-segmentation*”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8, Anchorage, AK, S.U.A., iunie 2008.

¹¹³ Bleyer, Michael, Rother, Carsten, Kohli, Pushmeet, „*Surface stereo with soft segmentation*”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1570–1577, San Francisco, CA, S.U.A., iunie 2010.

¹¹⁴ Lin, Michael H., Tomasi, Carlo, „*Surfaces with occlusions from layered stereo*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(8):1073–1078, 2003.

¹¹⁵ *Idem* 112.

5.1 Segmentarea imaginilor

O partiționare a imaginii este obținută prin utilizarea unui algoritm de segmentare în regiuni de culoare omogenă [Gonzalez, 2004]¹¹⁶ [Trémeau, 2004]¹¹⁷. Se disting două abordări:

- **Abordarea „contur”** – Granițele dintre zonele de schimbare bruscă a valorilor pixelilor, adică contururile din imagine, sunt detectate și apoi utilizate pentru a delimita diferitele regiuni.
- **Abordarea „regiune”** – Pixelii vecini care se aseamănă sunt regrupați pentru a forma regiuni.

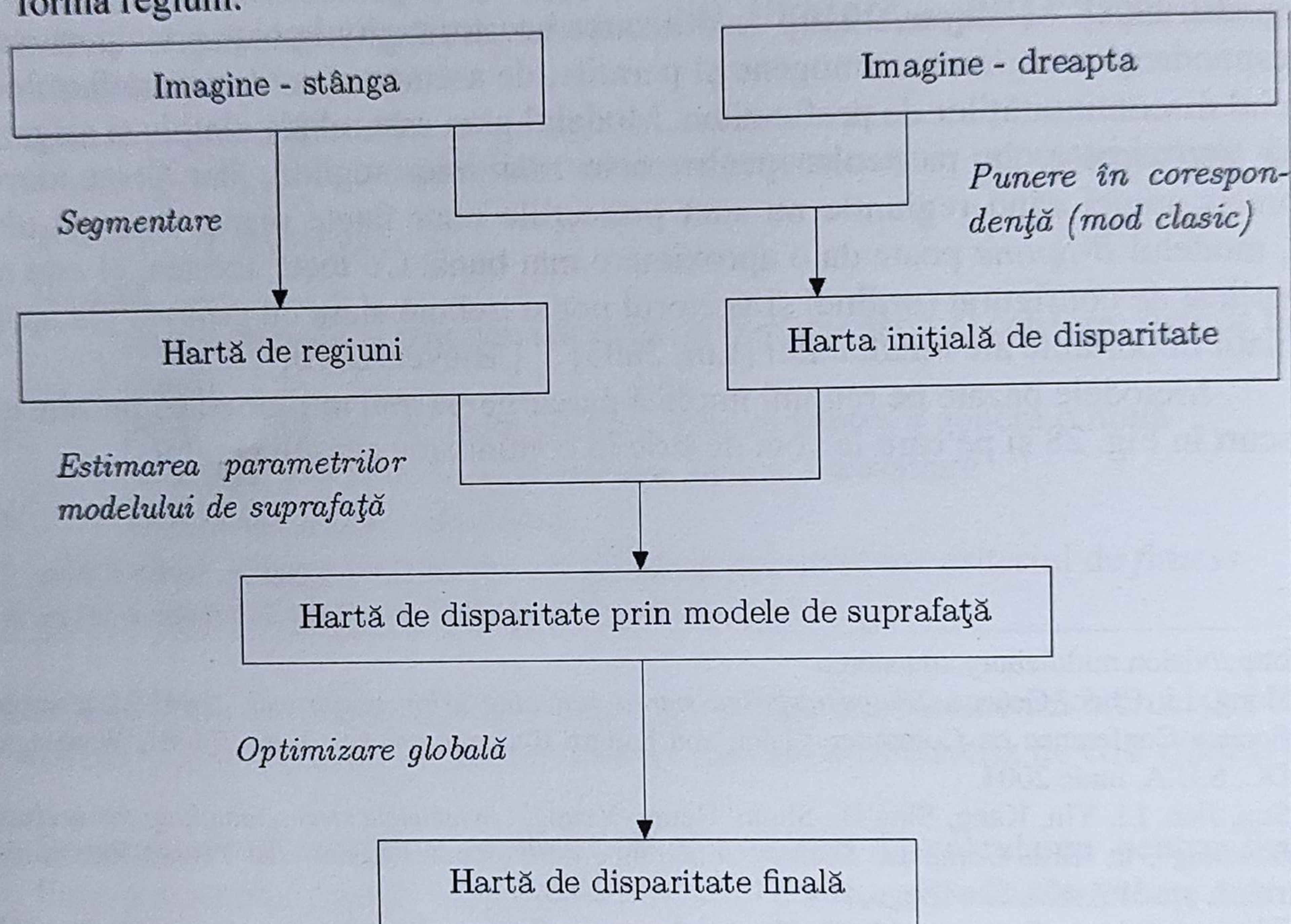


Fig. 28. Etapele metodelor de punere în corespondență a pixelilor bazate pe regiuni

Metoda *mean-shift* este cea mai utilizată în contextul punerii în corespondență a pixelilor pe bază de regiuni [Sun, 2002]¹¹⁸ [Bleyer, 2005]¹¹⁹ [Klaus, 2006]¹²⁰

¹¹⁶ Gonzalez, Rafael C., Woods, Richard E., Eddins, Steven L., „Digital image processing using Matlab”, Pearson Prentice Hall, 2004.

¹¹⁷ Trémeau, Alain, Fernandez-Maloigne, Christine, Bonton, Pierre, „Image numérique couleur – De l’acquisition au traitement”, Dunod, 2004.

¹¹⁸ Sun, Jian, Shum, Heung-Yeung, Zheng, Nan-Ning, „Stereo matching using belief propagation”, în European Conference on Computer Vision, pp. 450–452, Copenhagen, Danemarca, mai 2002.

¹¹⁹ Bleyer, Michaelm, Gelautz, Margrit, „A layered stereo matching algorithm using image segmentation and global visibility constraints”, ISPRS Journal of Photogrammetry and Remote Sensing, 59(3):128–150, mai 2005.

¹²⁰ Klaus, Aandreas, Sormann, Mario, Karner, Konrad, „Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure”, în International Conference on Pattern Recognition, vol. 3, pp. 15–18, Hong Kong, august 2006.

[Wang, 2008]¹²¹ [Yang, 2009]¹²² [Bleyer, 2010]¹²³. Spre deosebire de majoritatea algoritmilor de segmentare, acesta nu necesită specificarea numărului de regiuni la începutul procesării. În particular, acest parametru ar fi dificil de determinat *a priori* pentru că depinde în special de scenă. În plus, aproape întotdeauna este efectuată o sub-segmentare. Aceasta permite formarea de regiuni mici care sunt mai simplu de abordat de către un model plan.

Mean-shift. Detaliem aici metoda de segmentare *mean-shift* propusă în [Comaniciu, 1997]¹²⁴, care este utilizată pe scară largă pentru probleme de punere în corespondență a pixelilor. Metoda prezintă avantajul că nu trebuie specificat numărul de regiuni ce trebuie găsite. Segmentarea este de fapt o problemă de clasificare, care constă în regruparea pixelilor vecini care au valori apropiate. În termeni statistici, problema poate fi văzută ca estimarea funcției de densitate a culorii pixelilor.

Densitatea de probabilitate. Pornind de la un eșantion, adică de la un sub-set de observații dintre toate cazurile posibile, este posibil să fie estimată legea de probabilitate în cazul discret sau funcția de probabilitate în cazul continuu. Această lege sau funcție arată pentru fiecare valoare posibilă a unei variabile aleatoare X probabilitatea de a obține acea valoare.

În cazul discret, probabilitatea de a obține fiecare valoare a lui X este estimată prin raportarea numărului de apariții ale acelei valori la numărul de observații efectuate.

În cazul continuu, considerăm intervalul $[a; b]$. O funcție de densitate este definită prin:

$$\begin{aligned} f: R &\rightarrow R \\ x &\propto f(x) \end{aligned} \tag{53}$$

cu

$$p(a \leq x \leq b) = \int_a^b f(x) dx \tag{54}$$

unde $f(x) \geq 0$ și $\int_{-\infty}^{+\infty} f(x) dx = 1$.

¹²¹ Wang, Zeng-Fu, Zheng, Zhi-Gang, „A region based stereo matching algorithm using cooperative optimization”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pag. 1–8, Anchorage, AK, S.U.A., iunie 2008.

¹²² Yang, Qingxiong, Wang, Ling, Yang, Ruigang, Stewenius, Henrik, Nistér, David, „Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(3):492–504, 2009.

¹²³ Bleyer, Michael, Rother, Carsten, Kohli, Pushmeet, „Surface stereo with soft segmentation”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1570–1577, San Francisco, CA, S.U.A., iunie 2010.

¹²⁴ Comaniciu, Dorin, Meer, Peter, „Robust analysis of feature spaces: color image segmentation”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 750–755, San Juan, Porto Rico, iunie 1997.

Estimarea densității de probabilitate. Pentru a estima o funcție de densitate pornind de la un eșantion, se disting două tipuri de metode:

- **Metodele parametrice** – dacă se presupune un model pentru funcția de densitate (de exemplu cel gaussian), problema devine una de estimare a parametrilor acelui model;
- **Metodele non-parametrice** – dacă nu se presupune nici un model, se utilizează metodele non-parametrice. Este cazul metodei *mean-shift* în care funcția de densitate este estimată utilizând o estimare la nivel de nucleu (kernel).

Pornind de la un eșantion de n observații $x_i, i = 1, \dots, n$, funcția de densitate estimată, notată $\hat{f}(x)$, este calculată pentru $x \in R$ uni-dimensional prin:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (55)$$

unde h este un parametru de netezire și K este nucleul utilizat (conform Fig. 29). Am luat în considerare nucleul lui Epanechnikov (pentru că el permite simplificarea problemei pentru a utiliza algoritmul *mean-shift* detaliat în continuare). Expresia sa uni-dimensională este următoarea:

$$K_E(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{dacă } |x| \leq 1 \\ 0 & \text{dacă nu.} \end{cases} \quad (56)$$

Modurile unei densități de probabilitate prin *mean-shift*. Modurile funcției de densitate, adică valorile lui x pentru care $\hat{f}(x)$ are un maxim local, corespund „centrelor” diferitelor clase căutate (conform Fig. 29). Calculul explicit al acestor moduri după calcularea funcției de densitate poate fi evitat prin utilizarea algoritmului *mean-shift*. Acest algoritm este iterativ. La fiecare iterație sunt efectuate două etape:

- **Calculul mediei (*mean*)** – Media este calculată într-o fereastră a cărei dimensiuni rămân a fi definite. Fie x_i cele n elemente incluse în fereastra centrată pe elementul x considerat:

$$\mu(x) = \frac{1}{n} \sum_{i=1}^n x - x_i \quad (57)$$

- **Translație (*shift*)** – Elementele incluse în fereastra considerată sunt translatate cu media $\mu(x)$.

După mai multe iterații, este apoi posibil să se demonstreze că fiecare element traslatat converge către modul cel mai apropiat al funcției de densitate [Comaniciu, 1997]¹²⁵.

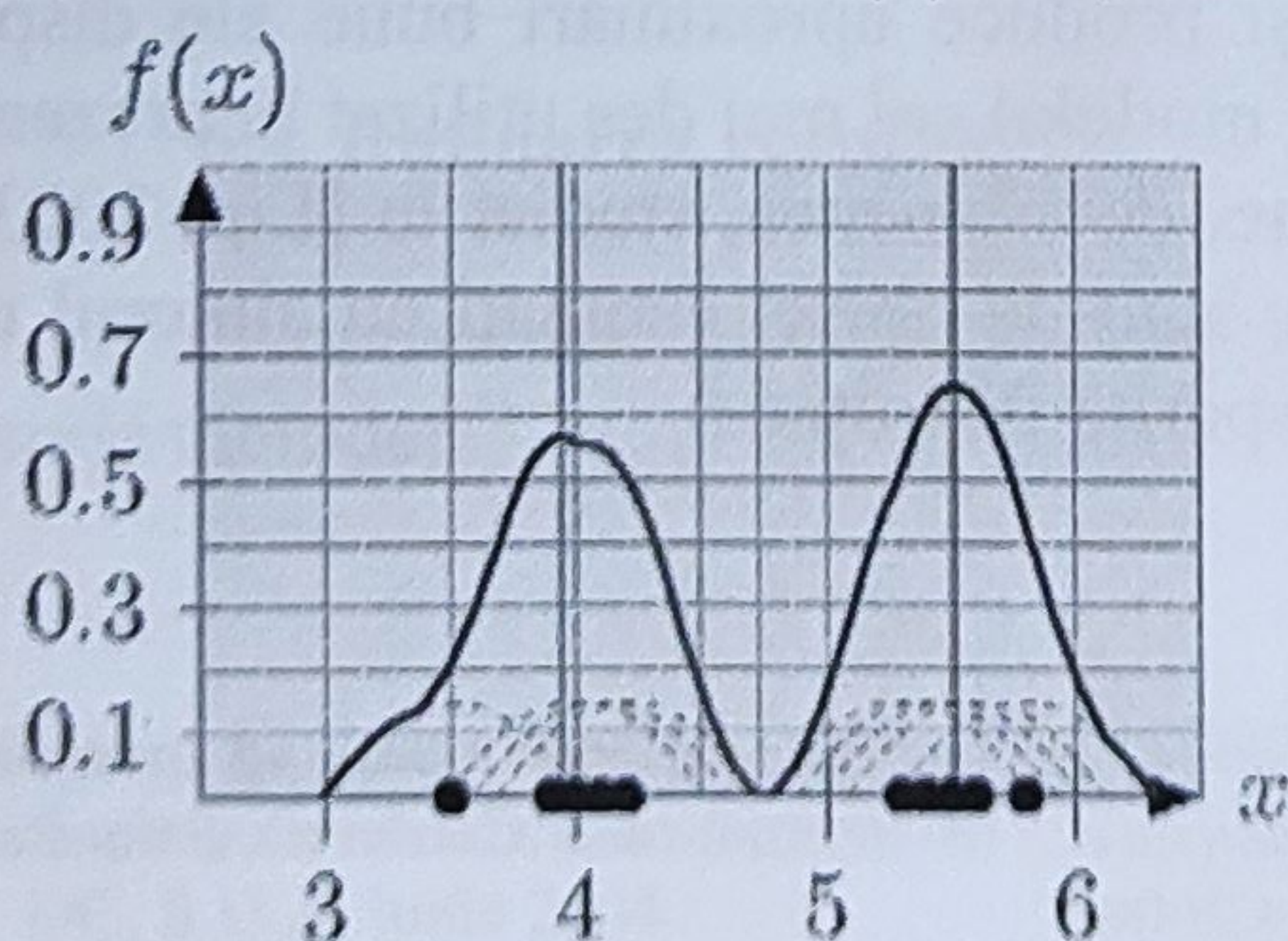
¹²⁵ Comaniciu, Dorin, Meer, Peter, „Robust analysis of feature spaces: color image segmentation”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 750–755, San Juan, Porto Rico, iunie 1997.

Analogia cu problema segmentării imaginii în regiuni. Problema segmentării imaginii în regiuni de culoare omogenă este similară celei de clasificare a elefanților din Fig. 29. Dar spațiul de clasificare este acum de trei sau cinci dimensiuni în loc să fie uni-dimensional: fiecare pixel din imagine reprezintă o observație pentru care luăm în considerare localizarea spațială, adică coordonatele pixelului din imagine, fie două dimensiuni, precum și nivelul de gri sau culoarea, fie, una sau trei dimensiuni suplimentare (conform Fig. 30). Aceste dimensiuni nu sunt deloc comparabile (în particular dimensiunile spațiale și componentele de culoare). Totuși, în practică, în loc să se recurgă la o normalizare, autorii lucrării [Comaniciu, 1997]¹²⁶ au propus utilizarea a doi parametri, notați h_s și h_r , pentru a defini fereastra centrată pe elementul considerat:

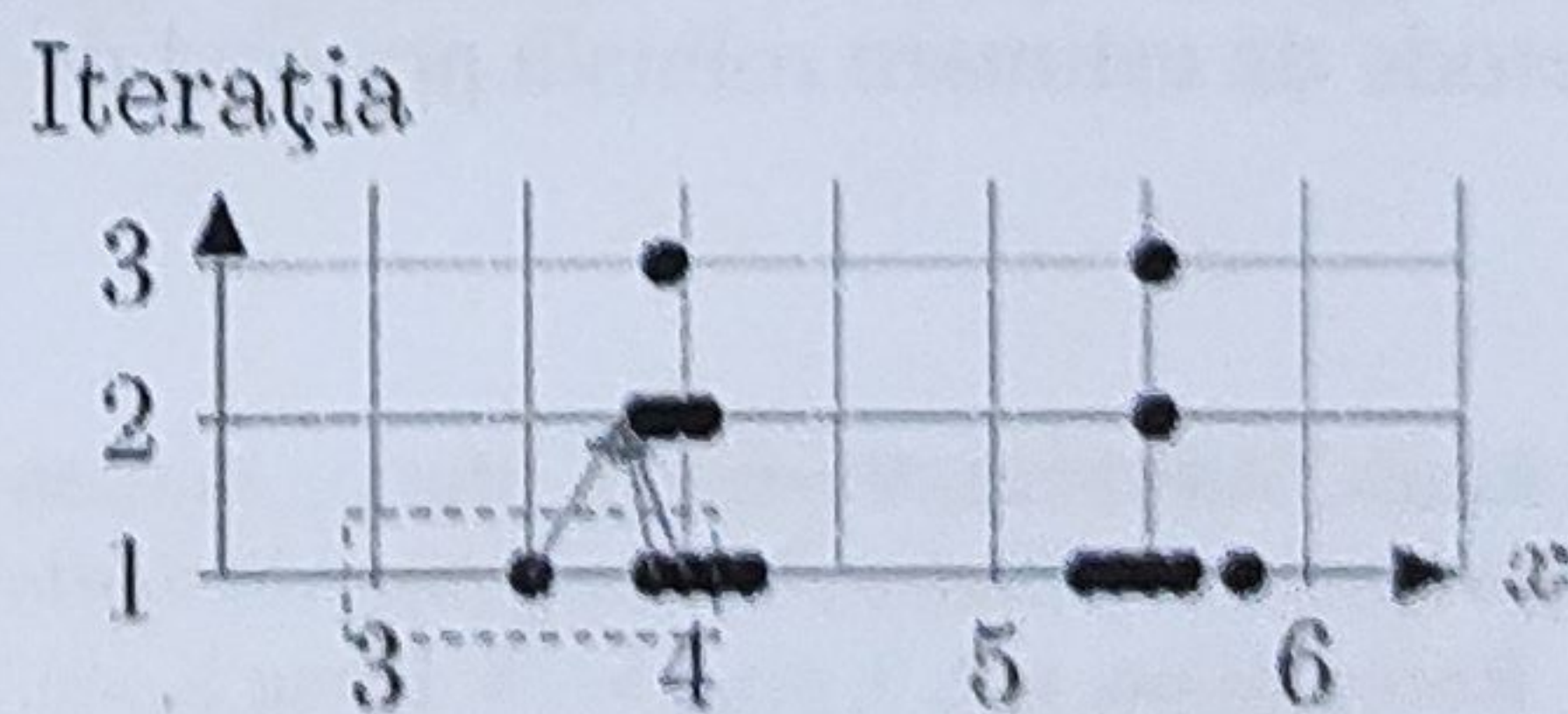
- h_s (*spatial*) – corespunde dimensiunii ferestrei conform dimensiunilor spațiale;
- h_r (*range*) – corespunde dimensiunii ferestrei conform dimensiunilor fotometrice (nivel de gri sau de culoare).

ELEFANT	x_i	$\hat{f}(x_i)$	CLASA
Bubble	4.0	0.55	C_1
Géant	4.2	0.51	C_1
Bobar	3.9	0.57	C_1
Colonel Défense	4.1	0.54	C_1
Tony	3.5	0.20	C_1
Norton	5.4	0.61	C_2
Palourde	5.8	0.42	C_2
Tantot	5.5	0.65	C_2
Yvoir	5.3	0.52	C_2
Muffin	5.6	0.62	C_2

(a) – Exemplu.



(b) – Funcția de densitate.



(c) – Mean-shift.

Fig. 29. Exemplu de segmentare prin mean-shift.

În figura 29, tabelul (a) prezintă greutatea x_i (în tone) măsurate pentru zece elefanți, precum și valorile funcției de densitate $\hat{f}(x)$ corespunzătoare. Aceste valori

¹²⁶ Idem.

sunt estimate prin metoda nucleului cu nucleu lui Epanechnikov ((b) cu linie continuă, funcția de densitate; nucleeele sunt punctate; valorile x_1 sunt marcate cu câte un punct negru pe axa absciselor). Modurile din $x_1 = 3,94$ și din $x_2 = 5,52$ sunt marcate cu o bară verticală. Astfel, elefanții a căror greutate este inferioară valorii de $3,94 + (5,52 - 3,94)/2 = 4,73$ tone (în verde în figura (b)) pot fi regrupați în clasa C_1 , iar cei a căror greutate este superioară acestei valori pot fi regrupați într-o clasă C_2 . Clasa fiecărui elefant este raportată în tabelul (a). Figura (c) demonstrează că putem regăsi aceste rezultate prin utilizarea algoritmului *mean-shift* în trei iterații. Luăm, de exemplu, primul punct negru din stânga. Fereastra utilizată este marcată punctat. Se calculează media elementelor incluse în această fereastră și acestea sunt deplasate cu această medie. Procesul este repetat de mai multe ori până la convergența spre un punct care corespunde modului cel mai apropiat al funcției de densitate. Considerăm apoi că punctele care au converș către un același mod aparțin unei aceleiași clase (sau regiuni).

5.2 Inițializarea disparităților

O metodă clasică de punere în corespondență a pixelilor este în general utilizată pentru această etapă. Pe de altă parte, tehnicile prezentate anterior (utilizarea condițiilor, adaptarea măsurii sau ferestrei de corelație la context, conform paragrafului 2) pot fi utilizate pentru limitarea numărului de erori [Klaus, 2006]¹²⁷ [Wang, 2008]¹²⁸ [Yang, 2009]¹²⁹.

5.3 Estimarea parametrilor modelului de suprafață

Modelul plan este relativ simplu și produce aproximări bune ale disparităților, în special pentru regiunile mici. Este modelul cel mai des utilizat în prezent în lucrările publicate, dar modelul *B-Spline* este, de asemenea, studiat în [Lin, 2003]¹³⁰ [Bleyer, 2010]¹³¹. Parametrii modelului ales sunt de obicei estimați cu ajutorul unei metode de estimare robustă pornind de la disparitățile inițiale:

¹²⁷ Klaus, Andreas, Sormann, Mario, Karner, Konrad, „Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure”, în International Conference on Pattern Recognition, vol. 3, pp. 15–18, Hong Kong, august 2006.

¹²⁸ Wang, Zeng-Fu, Zheng, Zhi-Gang, „A region based stereo matching algorithm using cooperative optimization”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8, Anchorage, AK, S.U.A., iunie 2008.

¹²⁹ Yang, Qingxiong, Wang, Ling, Yang, Ruigang, Stewénius, Henrik, Nistér, David, „Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(3):492–504, 2009.

¹³⁰ Lin, Michael H., Tomasi, Carlo, „Surfaces with occlusions from layered stereo”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(8):1073–1078, 2003.

¹³¹ Bleyer, Michael, Rother, Carsten, Kohli, Pushmeet, „Surface stereo with soft segmentation”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1570–1577, San Francisco, CA, S.U.A., iunie 2010.

- *Metoda celor mai mici pătrate aplicată iterativ* – Această tehnică este utilizată în special în [Hong, 2004]¹³². La fiecare iterație, disparitățile inițiale prea îndepărtate de modelul estimate sunt considerate ca date aberante și sunt eliminate pentru a nu influența rezultatul.
- *RANSAC* – Algoritmul RANSAC (*random sample consensus*) este utilizat în [Yang, 2008]¹³³ [Yang, 2009]¹³⁴.
- *Descompunerea parametrilor* – Această tehnică este utilizată în [Wang, 2008]¹³⁵ [Klaus, 2006]¹³⁶ pentru determinarea parametrilor unui plan. Prin scrierea funcției planului ca și o funcție de i și de j :

$$d_p : R^2 \rightarrow R$$

$$\begin{bmatrix} i \\ j \end{bmatrix} \alpha \quad p_1 i + p_2 j + p_3 = d_{i,j}$$

(58)

unde p_n (cu $n = 1, \dots, 3$) sunt parametrii căutați. Autorii remarcă faptul că:

$$p_1 = \frac{\partial d_p([i \ j]^T)}{\partial i} \quad p_2 = \frac{\partial d_p([i \ j]^T)}{\partial j} \quad p_3 = d_{i,j} - p_1 i - p_2 j. \quad (59)$$

Astfel, p_1 (respectiv p_2), sunt estimați luând toate perechile de pixeli posibile urmărind liniile (respectiv coloanele) pentru a estima derivatele, fiecare pereche producând o estimare pentru parametrul considerat. Se construiește apoi o histogramă a acestor estimări, care după aceea este netezită cu un nucleu gaussian. Valoarea care corespunde maximului acestei histograme este valoarea căutată pentru parametru. În sfârșit, un proces asemănător este utilizat pentru găsirea p_3 pornind de la setul de disparități inițiale ale regiunii.

5.4 Rafinarea rezultatelor

O optimizare globală permite rafinarea rezultatelor. Se pot distinge următoarele metode:

- ¹³² Hong, Li, Chen, George, „Segment-based stereo matching using graph cuts”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 74–81, Washington, DC, S.U.A., iunie 2004.
- ¹³³ Yang, Qingxiong, Engels, Chris, Akbarzadeh, Amir, „Near real-time stereo for weakly-textured scenes”, în British Machine Vision Conference, pp. 80–87, Leeds, U.K., septembrie 2008.
- ¹³⁴ Yang, Qingxiong, Wang, Ling, Yang, Ruigang, Stewenius, Henrik, Nistér, David, „Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(3):492–504, 2009.
- ¹³⁵ Wang, Zeng-Fu, Zheng, Zhi-Gang, „A region based stereo matching algorithm using cooperative optimization”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8, Anchorage, AK, S.U.A., iunie 2008.
- ¹³⁶ Klaus, Andreas, Sormann, Mario, Karner, Konrad, „Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure”, în International Conference on Pattern Recognition, vol. 3, pp. 15–18, Hong Kong, august 2006.

- *Atribuirea unui plan* – Diferitele plane de disparitate sunt obținute pornind de la modele estimate și fiecare pixel este atribuit unuia dintre aceste planuri în așa fel ca un cost global asupra întregii imagini să fie optimizat. Aceasta permite corectarea anumitor erori provenite în special dintr-o segmentare necorespunzătoare. Această soluție este utilizată în [Hong, 2004]¹³⁷ [Bleyer, 2010]¹³⁸ unde tehnica de optimizare folosită este tăietura grafului și în [Klaus, 2006]¹³⁹ unde tehnica de optimizare folosită este propagarea încrederii.
- *Rafinarea parametrilor* – Parametrii modelelor, estimați în prealabil, pot fi rafinați prin minimizarea unui cost global. Această soluție este în special utilizată în [Wang, 2008]¹⁴⁰ [Yang, 2008]¹⁴¹.

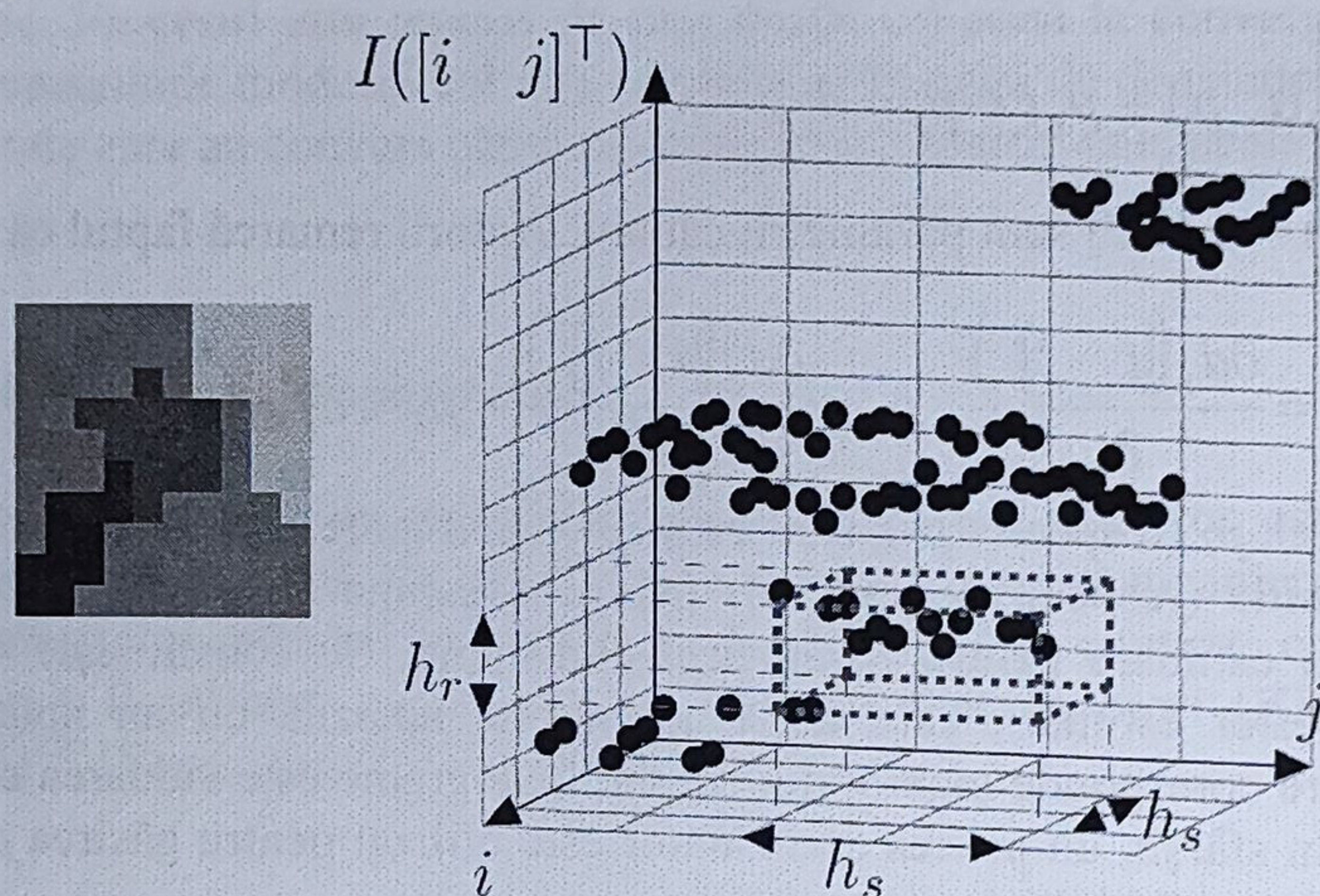


Fig. 30. *Spațiul de clasificare 3D* – Fiecare punct negru reprezintă un pixel caracterizat prin coordonatele sale din imagine (i și j) precum și prin nivelul său de gri $I([i \ j]^T)$

În general este utilizată culoarea, dar am utilizat nivelele de gri pentru a simplifica această ilustrație. Fereastra *mean-shift* de dimensiuni (h_s, h_s, h_r) este reprezentată punctat.

- ¹³⁷ Hong, Li, Chen, George, „Segment-based stereo matching using graph cuts”, în IEEE Computer Society Conf. on Computer Vision and Pattern Recogn, vol. 1, pp. 74–81, WA, DC, S.U.A, 2004.
- ¹³⁸ Bleyer, Michael, Rother, Carsten, Kohli, Pushmeet, „Surface stereo with soft segmentation”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1570–1577, San Francisco, CA, S.U.A., iunie 2010.
- ¹³⁹ Klaus, Andreas, Sormann, Mario, Karner, Konrad, „Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure”, în International Conference on Pattern Recognition, vol. 3, pp. 15–18, Hong Kong, august 2006.
- ¹⁴⁰ Wang, Zeng-Fu, Zheng, Zhi-Gang, „A region based stereo matching algorithm using cooperative optimization”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8, Anchorage, AK, S.U.A., iunie 2008.
- ¹⁴¹ Yang, Qingxiong, Engels, Chris, Akbarzadeh, Amir, „Near real-time stereo for weakly-textured scenes”, în British Machine Vision Conference, pp. 80–87, Leeds, U.K, septembrie 2008.

5.5 Punerea în corespondență a pixelilor prin sondajul regiunilor

Punerea în corespondență a pixelilor prin propagare dă rezultate bune când se caută un compromis între rata de acceptare¹⁴² și densitate. Cu cât rata de acceptare este mai ridicată, cu atât mai puțin dense sunt rezultatele. În [Gales, 2010]¹⁴³ [Gales, 2011]¹⁴⁴, se propune o abordare bazată pe un sondaj aleatoriu al regiunilor, cu scopul de a obține rate de acceptare ridicate dar și rezultate dense.

Principiul se bazează pe un proces aleator și iterativ care calculează pentru fiecare pixel, pornind de la mai multe segmentări, de la o hartă inițială de disparități și de la un model de suprafață, aproximări diferite ale valorii sale de disparitate. Fiecare aproximare constituie un vot și valoarea care obține cele mai multe sufragii este selecționată ca și disparitate finală. Această tehnică permite obținerea de rezultate satisfăcătoare fără să se recurgă la o etapă de rafinare prin optimizare globală.

Abordarea propusă este prezentată pe scurt în Fig. 31. Diferitele etape sunt:

- **Segmentări** – Fiecare regiune de culoare omogenă este presupusă a fi proiecția unei aceleiași suprafețe. Pixelii unei aceleiași regiuni au deci disparități apropiate. Metodele bazate pe regiuni utilizează în general o sub-segmentare, adică regiuni mici. De fapt, cu cât regiunile sunt mai mici, cu atât ele pot fi approximate cu un plan fără să se introducă o eroare mare. Cu toate acestea, cu cât regiunile sunt mai mici, cu atât este mai dificil să fie obținute disparitățile inițiale corecte în zonele apropiate de acoperiri sau de discontinuități de profunzime. Metoda propusă utilizează diferite niveluri de segmentare. Segmentările mai fine permit să se obțină valori de disparitate approximate mai precise în regiunile în care pixelii au disparități inițiale calculate de o manieră acceptabilă. Segmentările grosiere permit obținerea de voturi în zonele dificile și acoperite. De fapt, pentru pixelii situați în aceste zone riscul este mai mare decât la ceilalți pixeli de a nu avea calculată disparitatea inițială sau de a avea una eronată. Totuși, cu cât segmentarea este mai grosieră, cu atât cresc șansele de a regrupa acești pixeli împreună cu cei situați în zone care nu prezintă nici o dificultate în particular și pentru care disparitatea inițială este acceptabilă. Nu e mai puțin adevărat că se crește de asemenea riscul de a regrupa pixeli care nu aparțin unei aceleiași suprafețe (dar acest fenomen este limitat de segmentările mai fine). Rezultatele experimentale descrise în [Gales, 2010]¹⁴⁵ [Gales, 2011]¹⁴⁶ indică faptul că 3 sau 4 nivele de segmentare sunt necesare.

¹⁴² Rata de acceptare este numărul de împerecheri acceptate raportat la numărul de împerecheri efectuate. O împerechere este acceptată dacă corespondentul găsit este situat la o distanță inferioară sau egală cu 1 pixel față de corespondentul teoretic.

¹⁴³ Gales, Guillaume, Crouzil, Alain, Chambon, Sylvie, „A region-based randomized voting scheme for stereo matching”, în International Conference on Visual Computing, volume 6454 de Lecture Notes in Computer Science, pp 182–191, novembre 2010.

¹⁴⁴ Gales, Guillaume, „Mise en correspondance de pixels pour la stéréovision binoculaire par propagation d'appariements de points d'intérêt et sondage de régions”, teză de doctorat, Universitatea din Toulouse, iulie 2011.

¹⁴⁵ Idem 142.

¹⁴⁶ Idem 143.

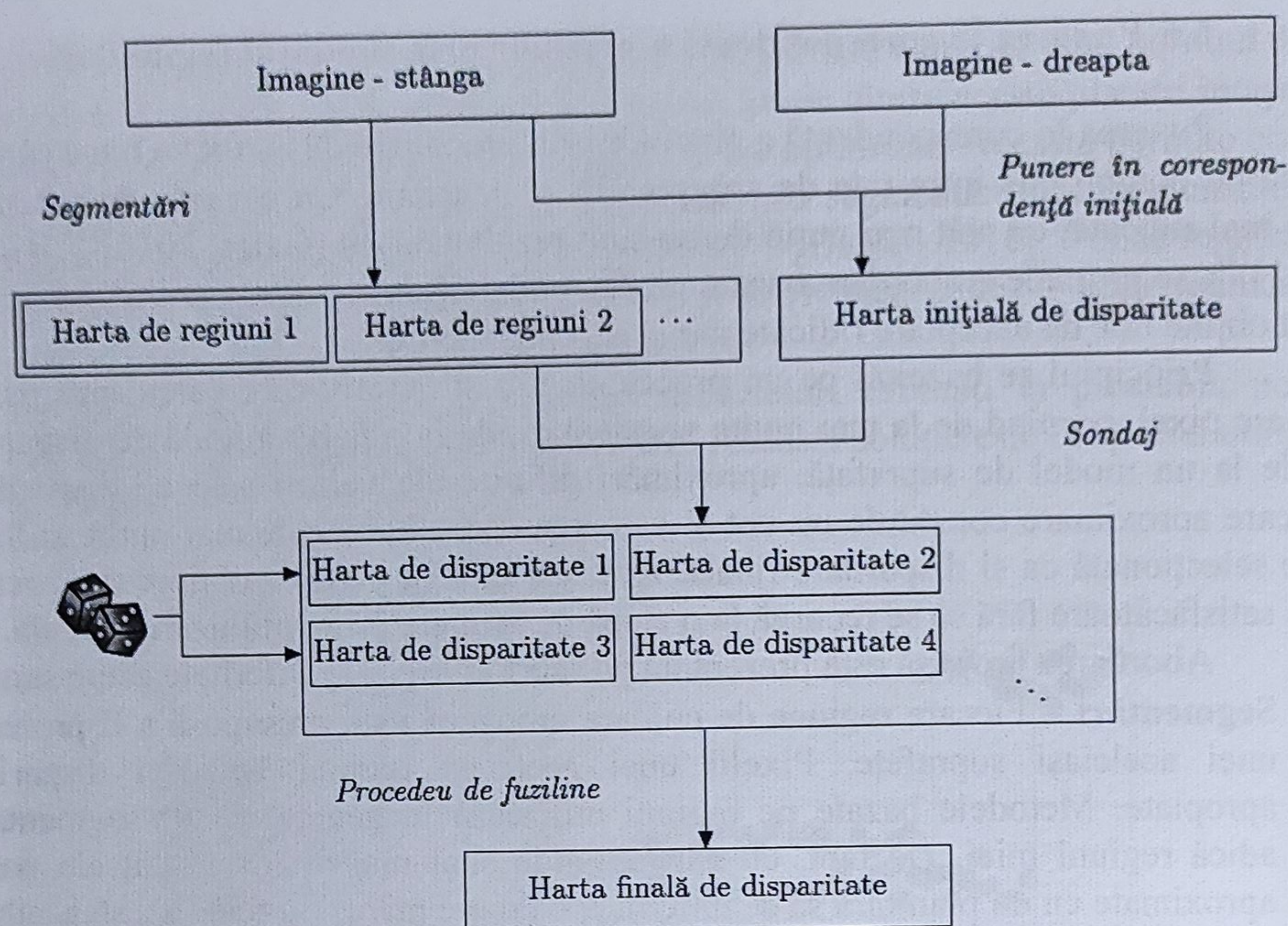


Fig. 31. Etapele punerii în corespondență a pixelilor prin sondarea regiunilor

- **Punerea inițială în corespondență a pixelilor** – Aceasta permite obținerea unei hărți de disparitate care este utilizată pentru calcularea parametrilor diferitelor plane pentru fiecare regiune. Această hartă trebuie să prezinte *a priori* un compromis bun între densitate și rata de acceptare. Densitatea ridicată permite mai multe disparități pentru sondaj și un număr mai important de voturi. O rată de acceptare ridicată permite reducerea numărului de voturi, conducând la rezultate eronate. În [Gales, 2010]¹⁴⁷ [Gales, 2011]¹⁴⁸ se arată că o punere în corespondență locală cu măsura CENSUS (conf. paragrafului 2) sau o propagare de germeni (conf. paragrafului 3) constituie o inițializare de calitate bună.
- **Sondaj** – Pentru fiecare regiune a fiecărui nivel de segmentare, o tripletă de puncte $\{[i_1 \ j_1 \ d_1]^T, [i_2 \ j_2 \ d_2]^T, [i_3 \ j_3 \ d_3]^T\}$ este trasă la sorți în spațiul disparităților inițiale (pixelii care nu au o disparitate inițială calculată nu sunt luați în calcul la tragerea la sorți). Parametrii planului, $[p_1 \ p_2 \ p_3]^T$, sunt apoi calculați ca:

¹⁴⁷ Gales, Guillaume, Crouzil, Alain, Chambon, Sylvie, „A region-based randomized voting scheme for stereo matching”, în International Conference on Visual Computing, volume 6454 de Lecture Notes in Computer Science, pp. 182–191, novembre 2010.

¹⁴⁸ Gales, Guillaume, „Mise en correspondance de pixels pour la stéréovision binoculaire par propagation d'appariements de points d'intérêt et sondage de régions”, teză de doctorat, Universitatea din Toulouse, iulie 2011.

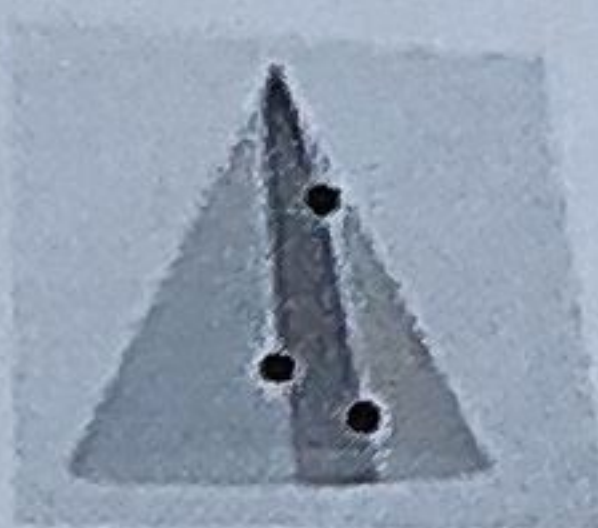
$$\begin{bmatrix} i_1 & j_1 & 1 \\ i_2 & j_2 & 1 \\ i_3 & j_3 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}. \quad (60)$$

TRAGERE LA SORTI

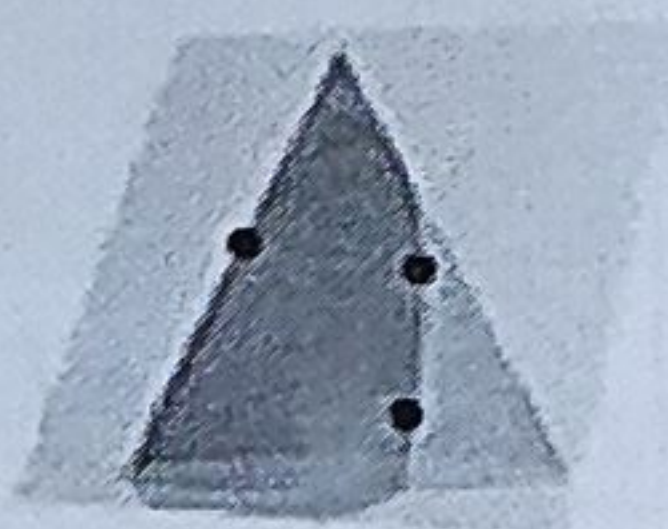
EXEMPLU

APROXIMAREA DISPARITĂȚILOR

1



2



...

...

...

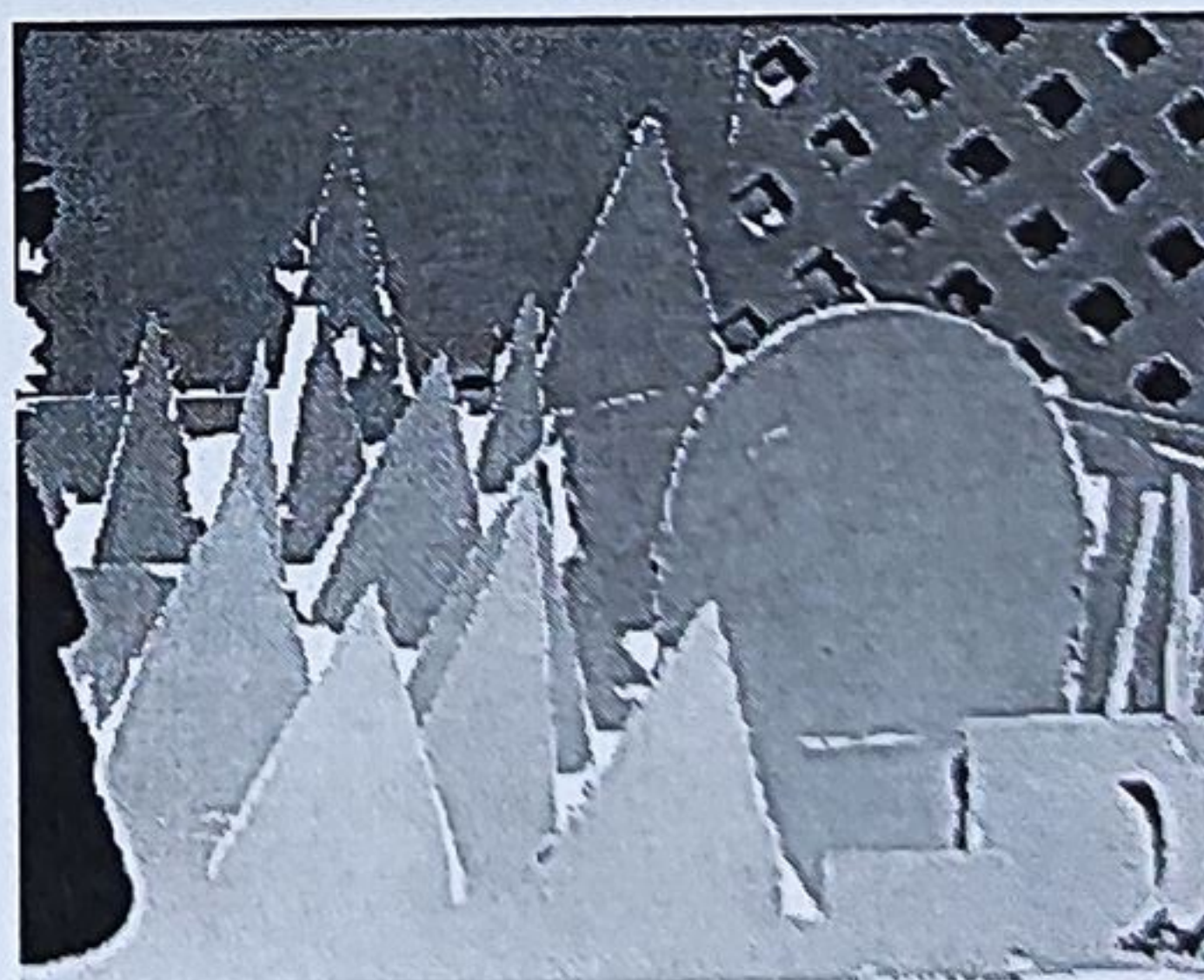


Fig. 32. Principiul sondajului regiunilor. În coloana centrală se prezintă planele induse (cu gri) de tripletele de puncte trase la sorti pentru o regiune reprezentând un con. Ultima coloană prezintă disparitățile aproximale induse de un triplet tras la sorti dintr-o regiune. Disparitățile eronate apar în culoare albă.

Principiul sondajului regiunilor este ilustrat în Fig. 32 printr-un exemplu de regiune reprezentând un con. La fiecare tragere, 3 puncte sunt trase la sorti și planul care trece prin aceste trei puncte permite calcularea unei disparități apropiate pentru fiecare pixel al regiunii. Harta de disparități obținută este prezentată în cea de-a treia coloană unde disparitățile eronate sunt marcate cu alb. Procesul este repetat de mai multe ori pentru fiecare regiune ale diferitelor segmentări. Se constată că la fiecare tragere, o mare parte din disparitățile approximate sunt corecte. Ideea este de a

combina diferitele hărți approximate pentru a obține harta finală de disparitate. Pe urmă, fiecare pixel aparținând regiunii primește un vot pentru disparitate, dat de:

$$v_n = p_1 k + p_2 l + p_3 \quad (61)$$

unde v_n este valoarea de disparitate a celui de-al n -lea vot pentru pixelul de coordonate $[k \ l]^T$. Procesul este repetat de m ori pentru a obține m voturi. Rezultatele raportate în [Gales, 2010]¹⁴⁹ [Gales, 2011]¹⁵⁰ arată că cel puțin 25 de trageri la sorți sunt necesare. Tragerea la sorți poate fi realizată într-o manieră:

- *Uniformă* – Toți tripleții au aceeași probabilitate de a fi trași la sorți.
- *Părtinitoare* – O valoare de încredere este calculată pentru fiecare triplet. Această valoare determină probabilitatea de acceptare a fiecărui punct. Aceasta permite limitarea numărului de selecții ale tripleților care conțin un punct a cărui disparitate inițială este eronată. Măsura de încredere, notată ct , a fiecărui pixel depinde de similaritatea fiecărui pixel și de cea a vecinilor săi cu pixelii lor corespondenți estimați:

$$ct : N^2 \rightarrow R$$

$$\begin{bmatrix} i \\ j \end{bmatrix} \propto \left(1 - \frac{\sum_{k=-r}^r \sum_{l=-r}^r \left| I^s \left(\begin{bmatrix} i+k \\ j+l \end{bmatrix} \right) - I^d \left(\begin{bmatrix} i+k \\ j+l+d_{i+k,j+l} \end{bmatrix} \right) \right|}{z} \right)^p \quad (62)$$

unde r stabilește dimensiunile vecinătății considerate, z este un termen de normalizare și p este o pondere. Cu cât p este mai ridicat, cu atât tragerea este mai părtinitoare în favoarea pixelilor care au valorile de confidență cele mai ridicate. Nu e mai puțin adevărat că, dacă p este prea ridicat, există tot timpul riscul de a trage la sorți aceiași tripleți. Rezultatele prezentate în [Gales, 2010]¹⁵¹ [Gales, 2011]¹⁵² arată că tragerea la sorți părtinitoare cu $p = 10$ dă cele mai bune rezultate.

- **Fuziune** – În sfârșit, pentru fiecare pixel se atribuie ca valoare de disparitate valoarea care a obținut cel mai multe voturi. Pentru a calcula această valoare, și pentru a avea o disparitate cu precizie de sub-pixel, ea este estimată prin exami-

¹⁴⁹ Gales, Guillaume, Crouzil, Alain, Chambon, Sylvie, „A region-based randomized voting scheme for stereo matching”, în International Conference on Visual Computing, volume 6454 de Lecture Notes in Computer Science, pp. 182–191, novembre 2010.

¹⁵⁰ Gales, Guillaume, „Mise en correspondance de pixels pour la stéréovision binoculaire par propagation d'appariements de points d'intérêt et sondage de régions”, teză de doctorat, Universitatea din Toulouse, iulie 2011.

¹⁵¹ Idem 148.

¹⁵² Idem 149.

narea modului principal al funcției de disparitate obținute prin estimare cu nucleul Epanechnikov. Este o tehnică de estimare prezentată în [Chen, 2002]¹⁵³:

$$d_{k,l} \leftarrow \arg \max_{v=1,K,d_{\max}} \frac{1}{m} \sum_{n=1}^m \begin{cases} \frac{3}{4} (1 - \|v - v_n\|^2) & \text{dacă } \|v - v_n\| \leq 1 \\ 0 & \text{dacă } nu. \end{cases} \quad (63)$$

Fuziunea este ilustrată pe ultimul rând din Fig. 32. Fig. 33 prezintă disparitățile finale obținute pentru perechea de imagini *Conuri* prin sondajul regiunilor.



Fig. 33. Harta finală de disparități a perechii de imagini *Conuri* obținute prin sondajul regiunilor

Concluzie

Obiectivul acestui capitol a fost de a descrie principalele metode de stabilire a corespondențelor între pixeli în cadrul stereoviziunii binoculare. Această etapă a făcut obiectul foarte multor lucrări de cercetare. Este o problemă pentru care este dificilă identificarea unei metode generice și satisfăcătoare pentru orice situație. Alegerea algoritmului depinde de mulți factori, cum ar fi proporția tolerabilă de erori, densitatea rezultatului și timpul de calcul. O aceeași metodă poate fi mai mult sau mai puțin performantă în funcție de tipul de scenă, adică de numărul de suprapuneri, de numărul de rupturi de profunzime sau de numărul de suprafețe puțin texturate. Calitatea rezultatelor poate deci varia de asemenea în interiorul aceleiași imagini, în funcție de zona considerată.

În fața abundenței de lucrări științifice, studiile comparative s-au dovedit necesare. Situl web „Middlebury Stereo Vision Page”¹⁵⁴ care însoțește articolul [Scharstein, 2002]¹⁵⁵ a fost o inițiativă foarte interesantă care a devenit o referință pentru un mare număr de cercetători din domeniu.

¹⁵³ Chen, Haifeng, Meer, Peter, „Robust computer vision through kernel density estimation”, în European Conference on Computer Vision, vol. 1, pp. 236–250, Copenhaga, Danemarca, mai–iunie 2002.

¹⁵⁴ <http://vision.middlebury.edu/stereo/>

¹⁵⁵ Scharstein, Daniel, Szeliski, Richard, „A taxonomy and evaluation of dense twoframe stereo correspondence algorithms”, International Journal of Computer Vision, 47(1):7–42, 2002.

Bibliografie

- Banks, Jasmine, Bennamoun, Mohammed, „*Reliability analysis of the rank transform for stereo matching*”, IEEE Transactions on Systems, Man, and Cybernetics, 31(6):870–880, 2001.
- Bay, Herbert, Ess, Andreas, Tuytelaars, Tinne, Gool, Luc Van, „*SURF: Speeded up robust features*”, Computer Vision and Image Understanding, 110(3):346–359, 2006.
- Bleyer, Michaelm, Gelautz, Margrit, „*A layered stereo matching algorithm using image segmentation and global visibility constraints*”, ISPRS Journal of Photogrammetry and Remote Sensing, 59(3):128–150, mai 2005.
- Bleyer, Michael, Rother, Carsten, Kohli, Pushmeet, „*Surface stereo with soft segmentation*”, în IEEE Computer Society Conf. on Computer Vision&Pattern Recognition, pp. 1570–1577, San Francisco, CA, USA, iunie 2010.
- Bobick, Aaron F., Intille, Stephen S., „*Large occlusion stereo*”, International Journal of Computer Vision, 33(3): pp. 181–200, 1999.
- Bocquillon, Benoît, „*Contributions à l'autocalibrage des caméras: modélisations et solutions garanties par l'analyse d'intervalle*”, teză de doctorat, Universitatea din Toulouse, Oct. 2008.
- Cech, Jan, „*Accurate and Robust Stereoscopic Matching in Efficient Algorithms*”, teză de doctorat, Universitatea Tehnică din Pragă, Cehia, februarie 2009.
- Chambon, Sylvie, Crouzil, Alain, „*Colour correlation-based matching*”, International Journal of Robotics and Automation, 20(2):78–87, 2005.
- Chambon, Sylvie, Crouzil, Alain, „*Similarity measures for image matching despite occlusions in stereo vision*”, Pattern Recognition, 44:2063–2075, 2011.
- Chambon, Sylvie, Crouzil, Alain, „*Combination of correlation measures for dense stereo matching*”, în International Conference on Computer Vision Theory and Applications, material electronic, Algarve, Portugal, martie 2011(b).
- Chen, Haifeng, Meer, Peter, „*Robust computer vision through kernel density estimation*”, în European Conference on Computer Vision, vol. 1, pp. 236–250, Copenhagen, Danemarca, mai–iunie 2002.
- Chen, Qian, Medioni, Gérard, „*A volumetric stereo matching method: Application to image-based modeling*”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pages 1029–1034, Fort Collins, SUA, iunie 1999.
- Comaniciu, Dorin, Meer, Peter, „*Robust analysis of feature spaces: color image segmentation*”, în IEEE Computer Society Conf. on Computer Vision&Pattern Recogn., pp. 750–755, San Juan, Porto Rico, juin 1997.
- Crouzil, Alain, Massip-Pailhes, Louis, Castan, Serge, „*A new correlation criterion based on gradient fields similarity*”, International Conference on Pattern Recognition, Vol. 1, pp. 632–636, Viena, Austria, august 1996.
- Devernay, Frederic, Faugeras, Olivier, „*Computing differential properties of 3-D shapes from stereoscopic images without 3-D models*”, raport de cercetare RR-2304, Institut National de Recherche en Informatique et en Automatique, INRIA, 1994.
- Ford, Lester R. Jr., Fulkerson, Delbert R., „*A simple algorithm for finding maximal network flows and an application to the hitchcock problem*”, Canadian Journal of Mathematics, 9:210–218, 1957.
- Freeman, William. T., Pasztor, Egon C., Carmichael, Owen T., „*Learning low-level vision*”, International Journal of Computer Vision, 40(1):25–47, 2000.
- Fua, Pascal, „*A parallel stereo algorithm that produces dense depth maps and preserves image features*”, Intern. Journal of Machine Vision and Applications, 6:35–49, 1993.
- Gales, Guillaume, Crouzil, Alain, Chambon, Sylvie, „*A region-based randomized voting scheme for stereo matching*”, în International Conference on Visual Computing, volume 6454 de Lecture Notes in Computer Science, pp. 182–191, novembre 2010.
- Gales, Guillaume, „*Mise en correspondance de pixels pour la stéréovision binoculaire par propagation d'appariements de points d'intérêt et sondage de régions*”, teză de doctorat, Univ. din Toulouse, iulie 2011.
- Gales, Guillaume, Chambon, Sylvie, Crouzil, Alain, McDonald, John, „*Reliability measure for propagation-based stereo matching*”, în Workshop on Image Analysis for Multimedia Interactive Services, Dublin City University, Irlanda, mai 2012.

Stabilirea corespondențelor între pixeli în stereoviziunea binoculară

- Garcia, Dorian, „*Mesures de formes et de champs de déplacements tridimensionnels par stereo-correlation d'images*”, teză de doctorat, Institut National Polytechnique de Toulouse, decembrie 2001.
- Geman, Stuart, Geman, Donald, „*Stochastic relaxation, Gibbs distribution, and the bayesian restoration of images*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6(6):721–741, 1984.
- Gong, Minglun, Yang, Yee-Hong, „*Fast stereo matching using reliability-based dynamic programming and consistency constraints*”, IEEE International Conference on Computer Vision, Nisa, Franța, Vol. 1, pp. 610–617, octombrie 2003.
- Gong, Minglun, Yang, Ruigang, Wang, Liang, Gong, Mingwei, „*A performance study on different cost aggregation approaches used in real-time stereo matching*”, International Journal of Computer Vision, Vol. 2, nr. 75, pp. 283–296, 2007.
- Gonzalez, Rafael C., Woods, Richard E., Eddins, Steven L., „*Digital image processing using Matlab*”, Pearson Prentice Hall, 2004.
- Gruen, Armin W., „*Adaptive least squares correlation: a powerful image matching technique*”, South African Journal of Photogrammetry, Remote Sensing and Cartography, vol. 3, nr. 14, pp. 175–187, 1985.
- Harris, Chris, Stephens, Mike, „*A combined corner and edge detector*”, în Alvey Vision Conference, pp. 147–151, Manchester, UK, ianuarie 1988.
- Hartley, Richard, Zisserman, Andrew, „*Multiple view geometry*”, Cambridge University Press, 2-nd ed., 2003.
- Harvent, Jacques, „*Mesures de formes par correlation multi-images: application a l'inspection de pieces aeronautiques a l'aide d'un systeme multi-cameras*”, teză de doctorat, Univ. din Toulouse, nov. 2010.
- Hirschmuller, Heiko, Innocent, Peter R., Garibaldi, Jon, „*Real-time correlation-based stereo vision with reduced border errors*”, International Journal of Computer Vision, 47(1-3): pp. 229–246, 2002.
- Holland, John H., „*Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence*”, MIT Press, 1975.
- Hong, Li, Chen, George, „*Segment-based stereo matching using graph cuts*”, în IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 74–81, Washington, DC, S.U.A., iunie 2004.
- Je, Frédéric, Schmid, Cordelia, „*Scale-invariant shape features for recognition of object categorie*”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 90–96, Washington, DC, S.U.A., iunie 2004.
- Kanade, Takeo, Okutomi, Masatoshi, „*A stereo matching algorithm with an adaptive window: Theory and experiment*”, IEEE Trans. on Pattern Analysis and Machine Intelligence, 16(9): pp. 920–932, 1994.
- Kannala, Juho, Brandt, Sami S., „*Quasi-dense wide baseline matching using meci propagation*”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8, USA iunie 2007.
- Klaus, Andreas, Sormann, Mario, Karner, Konrad, „*Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure*”, în International Conference on Pattern Recognition, vol. 3, pp. 15–18, Hong Kong, august 2006.
- Kostkrova, Jana, Sara, Radim, „*Disparity component matching revisited*”, Raport tehnic CTU-CMP-2002-08, Universitatea Tehnică din Praga, 2002.
- Lhuillier, Maxime, Quan, Long, „*Match propagation for image-based modeling and rendering*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(8): 1140–1146, 2002.
- Li, Stan Z., „*Markov random field models in computer vision*”, în European Conference on Computer Vision, vol. 2, pp. 361–370, Stockholm, Suedia, mai 1994.
- Lin, Michael H., Tomasi, Carlo, „*Surfaces with occlusions from layered stereo*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(8):1073–1078, 2003.
- Lindeberg, Tony, „*Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention*”, International Journal of Computer Vision, 1(3):1074–1130, 1993.
- Loop, Charles T., Zhang, Zhengyou, „*Computing rectifying homographies for stereo vision*”, in IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 125–131, Fort Collins, USA June 1999.

- Lotti, Jean-Luc, Giraudon, Gerard, „Correlation algorithm with adaptive window for aerial image in stereo vision”, în European Symposium on Satellite Remote Sensing, pp. 2315-2325, Roma, septembrie 1994.
- Lowe, David G., „Object recognition from local scale-invariant features”, în IEEE International Conference on Computer Vision, vol. 2, pag. 1150–1157, Corfu, Grecia, septembrie 1999.
- Mikolajczyk, Krystian, „Detection of local features invariant to affine transformations. Application to matching and recognition”, Teză de doctorat, Institut National Polytechnique de Grenoble, iunie 2002.
- Mikolajczyk, Krystian, Schmid, Cordelia, „Scale & affine invariant interest point detectors”, International Journal of Computer Vision, 60(1):63–86, 2004.
- Mikolajczyk, Krystian, Schmid, Cordelia, „A performance evaluation of local descriptors”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(10):1615–1630, 2005.
- Moravec, Hans.P., „Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover”, Teză de doctorat, Universitatea Stanford, Stanford, California, S.U.A., septembrie 1980.
- Otto, G. Paul, Chau, Tony K. W., „„Region-growing” algorithm for matching of terrain images”, Image and Vision Computing, 7(2):83–94, 1989.
- Parisot, Pascaline, „Suivi d'objets dans les séquences d'images de scènes déformables: de l'importance des points d'intérêt et du maillage 2D”, Teză de doctorat, Universitatea din Toulouse, ianuarie 2009.
- Park, Joonyoung, Kim, Wonsik, Lee, Kyoung M., „Stereo matching using population based MCMC”, în Asian Conference on Computer Vision, vol. 2, pp. 560–569, Tokyo, Japonia, noiembrie 2007.
- Pearl, Judea, „Reverend Bayes on inference engines: A distributed hierarchical approach”, în National Conference on Artificial Intelligence, pp. 133–136, Pittsburgh, PA, S.U.A., august 1982.
- Rosten, Edward, Drummond, Tom, „Machine learning for high-speed corner detection”, în European Conference on Computer Vision, vol. 1, pp. 430–443, Graz, Austria, mai 2006.
- Rousseeuw, Peter J., Croux, Christophe, „ L_1 -statistical analysis and related methods”, în Y. Dodge, editor, Explicit Scale Estimators with High Breakdown point, pp. 77-92, Elsevier, Amsterdam, Olanda, 1992.
- Roy, Sébastien, „Stereo without epipolar lines: A maximum-flow formulation”, International Journal of Computer Vision, 34(2–3):147–161, 1999.
- Scharstein, Daniel, Szeliski, Richard, „A taxonomy and evaluation of dense twoframe stereo correspondence algorithms”, International Journal of Computer Vision, 47(1):7–42, 2002.
- Schmid, Cordelia, Mohr, Roger, Bauckhage, Christian, „Evaluation of interest point detectors”, International Journal of Computer Vision, 37(2):151-172, 2000.
- Smith, Stephen M., Brady, J. Michael, „SUSAN – a new approach to low level image processing”, International Journal of Computer Vision, 23(1):45–78, 1997.
- Szeliski, Richard, „Bayesian modeling of uncertainty in low-level vision”, International Journal of Computer Vision, 5(3):271–302, decembrie 1990.
- Szeliski, Richard, „Computer Vision: Algorithms and Applications”, Springer, 2010.
- Sun, Jian, Shum, Heung-Yeung, Zheng, Nan-Ning, „Stereo matching using belief propagation”, în European Conference on Computer Vision, pp. 450–452, Copenhaga, Danemarca, mai 2002.
- Sun, Jian, Li, Yin, Kang, Sing B., Shum, Heung-Yeung, „Symmetric stereo matching for occlusion handling”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 399–406, San Diego, CA, S.U.A., iunie 2005.
- Sur, Frédéric, Noury, Nicolas, Berger, Marie-Odile, „Determining point correspondences between two views under geometric constraint and photometric consistency”, Raport de cercetare RR-7246, Institut National de Recherche en Informatique et en Automatique, INRIA, 2010.
- Taguchi, Yuichi, Wiburn, Bennett, Zitnick, C. Lawrence, „Stereo reconstruction with mixed pixels using adaptive over-segmentation”, în IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1–8, Anchorage, AK, S.U.A., iunie 2008.
- Trémeau, Alain, Fernandez-Maloigne, Christine, Bonton, Pierre, „Image numérique couleur – De l'acquisition au traitement”, Dunod, 2004.
- Tuytelaars, Tinne, Mikolajczyk, Krystian, „Local invariant feature detectors: a survey”, Foundations and Trends® in Computer Graphics and Vision, 3(3):177–280, 2008.
- Veksler, Olga, „Efficient Graph-Based Energy Minimization Methods in Computer Vision”, teză de doctorat, Cornell University, Ithaca, NY, S.U.A., august 1999.

Stabilirea corespondențelor între pixeli în stereoviziunea binoculară

- Vincent, Étienne, Laganière, Robert, „*Matching feature points in stereo pairs: A comparative study of some matching strategies*”, *Machine Graphics & Vision*, 10(3):237–259, 2001.
- Wang, Zeng-Fu, Zheng, Zhi-Gang, „*A region based stereo matching algorithm using cooperative optimization*”, în *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1–8, Anchorage, AK, S.U.A., iunie 2008.
- Xu, Li, Jia, Jiaya, „*Stereo matching: an outlier confidence approach*”, în *European Conference on Computer Vision*, vol. 4, pp. 775–787, Marsilia, Franța, octombrie 2008.
- Yang, Qingxiong, Engels, Chris, Akbarzadeh, Amir, „*Near real-time stereo for weakly-textured scenes*”, în *British Machine Vision Conference*, pp. 80–87, Leeds, U.K, septembrie 2008.
- Yang, Qingxiong, Wang, Ling, Yang, Ruigang, Stewenius, Henrik, Nistér, David, „*Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling*”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):492–504, 2009.
- Yoon, Kuk-Jin, Kweon, In-So, „*Locally adaptive support-weight approach for visual correspondence search*”, *IEEE Computer Society conference on Computer Vision and pattern recognition*, vol. 2, pp. 924–931, San Diego, California, Statele Unite, iunie 2005.
- Zabih, Ramin, Woodfill, John, „*Non-parametric local transforms for computing visual correspondence*”, *European Conference on Computer Vision*, pp. 151–158, Stockholm, Suedia, mai 1994.
- Zitnick, C. Lawrence, Kang, Sing B., „*Stereo for image-based rendering using oversegmentation*”, *International Journal of Computer Vision*, 75(1):49–65, 2007.
- Zhou, Ji, Shi, Jiaoying, „*A robust algorithm for feature point matching*”, *Computer and Graphics*, 26(3):429–436, 2002.
- Zhuang, Yong, Ma, Ma, Jun, Wei, Liu, Liu, „*Progressive matching based on segmentation for 3D reconstruction*”, în *International Conference on Computer and Information Technology*, pp. 575–579, Shanghai, China, septembrie 2005.

Tehnici de reconstrucție 3D din imagini multiple

FLORIN ROTARU, SILVIU BEJINARIU, SIMONA PESCARU¹

ABSTRACT. First part of this chapter introduces the main 3D reconstruction techniques from multiple views of moving objects. There are presented affine reconstruction from orthographic projection methods and projective reconstruction techniques. While the analysed projective reconstruction technique is restricted to a rigid moving object the affine reconstruction part is more elaborated. There are discussed reconstruction techniques first for one rigid object then for two or more independently moving rigid objects and finally for non rigid moving objects. In second part there is analysed a 3D tracking system of a mobile robot. The tracking is done in more successive steps: 1) feature extraction from current image; 2) 3D coordinates calculus by stereovision techniques of the extracted features; 3) 3D feature tracking in successive views; 4) current robot pose computation, expressed in a world coordinate system; 5) robot pose and 3D feature coordinates correction.

KEYWORDS: *multiple view reconstruction, affine projective transform, 3D tracking.*

1. Introducere

În prima parte a capitolului sunt analizate tehnici de reconstrucție din imagini multiple ale unor obiecte în mișcare. Sunt discutate tehnici de reconstrucție afină din imagini obținute în principal prin proiecție ortografică și tehnici de reconstrucție proiectivă.

În partea a doua sunt analizate sisteme de urmărire 3D a unui robot mobil. Urmărirea implică operațiile: extragerea din imaginea curentă a trăsăturilor care vor fi urmărite și în următoarele cadre, determinarea prin tehnici stereo a coordonatelor 3D ale trăsăturilor extrase, urmărirea trăsăturilor 3D în secvența de imagini, calculul pozițiilor robotului într-un sistem de referință extern iar în final corecția poziției robotului și a coordonatelor 3D ale caracteristicilor scenei.

2. Tehnici de reconstrucție afină și proiectivă din imagini multiple

2.1. Reconstrucție afină

Vom începe cu o teoremă care demonstrează că o reconstrucție afină poate fi obținută din două proiecții ortografice, după care vom studia problema reconstrucției afine pornind de la setul de n corespondențe din m imagini.

¹ Florin Rotaru, Silviu Bejinariu, cercetători științifici, IIT, Academia Română, Iași, Simona Pescaru, Chief Vision Scientist, Robotic Vision Technologies, Dynamic Ventures, Brașov, Romania.

Teorema 1: Din două reprezentări ortografice pentru patru puncte necoplanare structura afină a scenei este determinată în mod unic.

Așa cum se arată în figura 1 din două imagini ortografice a cinci caracteristici punctuale A, B, C, D și P se pot reconstrui coordonatele afine ale punctului P în baza (A, B, C, D).

Koenderink și Van Doorn au demonstrat² că acest lucru este posibil, plecând de la faptul că proiecția ortografică a unui plan pe un alt plan este o transformare afină. În particular, când punctul P aparține planului Π din figura 1 care conține triunghiul ABC, coordonatele sale afine în planul Π , format de aceste trei puncte, pot fi măsurate direct în oricare dintre cele două imagini Π' , respectiv Π'' .

Fie E (respectiv Q) intersecția dreptei care trece prin punctele D și d' (respectiv P și p') cu planul Π (figura 1). Proiecțiile e'' și q'' ale punctelor E și P pe planul Π'' au aceleași coordonate afine în baza (a'', b'', c'') ca și punctele d' și p' în baza (a', b', c') . În plus, deoarece segmentele ED și QP sunt paralele cu prima direcție a proiecției, segmentele $e''d''$ și $q''p''$ sunt de asemenea paralele. Se poate măsura raportul:

$$\lambda = \frac{\overline{q''p''}}{\overline{e''d''}} = \frac{\overline{QP}}{\overline{ED}} \quad (1)$$

unde \overline{AB} reprezintă distanța dintre punctele A și B pentru o orientare arbitrară a dreptei determinată de cele două puncte.

Se notează cu $(\alpha_{d'}, \beta_{d'})$ și $(\alpha_{p'}, \beta_{p'})$ coordonatele punctelor $d' = e'$ și $p' = q'$ în baza (a', b', c') . Avem:

$$\begin{aligned} \overrightarrow{AP} &= \overrightarrow{AQ} + \overrightarrow{QP} \\ &= \alpha_{p'} \overrightarrow{AB} + \beta_{p'} \overrightarrow{AC} + \lambda \overrightarrow{ED} \\ &= (\alpha_{p'} - \lambda \alpha_{d'}) \overrightarrow{AB} + (\beta_{p'} - \lambda \beta_{d'}) \overrightarrow{AC} + \lambda \overrightarrow{AD} \end{aligned} \quad (2)$$

Cu alte cuvinte, coordonatele afine ale punctului P în baza (A, B, C, D) sunt $(\alpha_{p'} - \lambda \alpha_{d'}, \beta_{p'} - \lambda \beta_{d'}, \lambda)$.

Folosind următoarele două teoreme referitoare la descompunerea SVD a matricelor vom analiza în continuare reconstrucția afină din m imagini pentru care s-au determinat m corespondențe a câte n pixeli reprezentând n puncte din spațiul afin.

Teorema 2: Când matricea A are rangul mai mare decât p, descompunerea SVD $U_p W_p V_p^t$ este cea mai bună aproximare a lui A de rang p (în sensul normei Frobenius).

Teorema 3: Fie $a_i \in \mathbb{R}^m$ ($i=1, n$) n vectori care reprezintă coloanele matricei A; subspațiul vectorial V_p de dimensiune p care minimizează eroarea pătratică medie,

² Jan J. Koendering, Andrea J. van Doorn, „Affine Structure from Motion”, în *Journal of the Optical Society of America A*, 8, pp. 377-385, 1991.

$$\frac{1}{n} \sum_{i=1}^n |a_i - b_i|^2 \quad (3)$$

unde b_i reprezintă proiecția ortogonală a lui a_i pe V_p , este subspațiul lui R^m generat de coloanele lui U_p .

O cameră afină având modelul $P_A = \begin{bmatrix} p_{11} & p_{12} & p_{13} & t_1 \\ p_{21} & p_{22} & p_{23} & t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ vizualizează punctele

$X_i = (X_i, Y_i, Z_i, 1)^t$ din planul proiectiv în pixelii $x_i = (x_i, y_i, 1)^t$ astfel încât relația $x_i = P_A X_i$ în coordonate omogene, exprimată în coordonate neomogene devine:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \quad (4)$$

unde s-au folosit notațiile în coordonate neomogene $x = (x, y)^t$ pentru punctele din planul imaginii, $X = (x, y, z)^t$ pentru punctele din spațiu, $t = (t_1, t_2)^t$ pentru vectorul translațiilor $(t_1, t_2, 1)^t$. De asemenea, notând cu P matricea 2×3 din (4), relația respectivă se rescrie ca:

$$x_i = P X_i + t \quad (5)$$

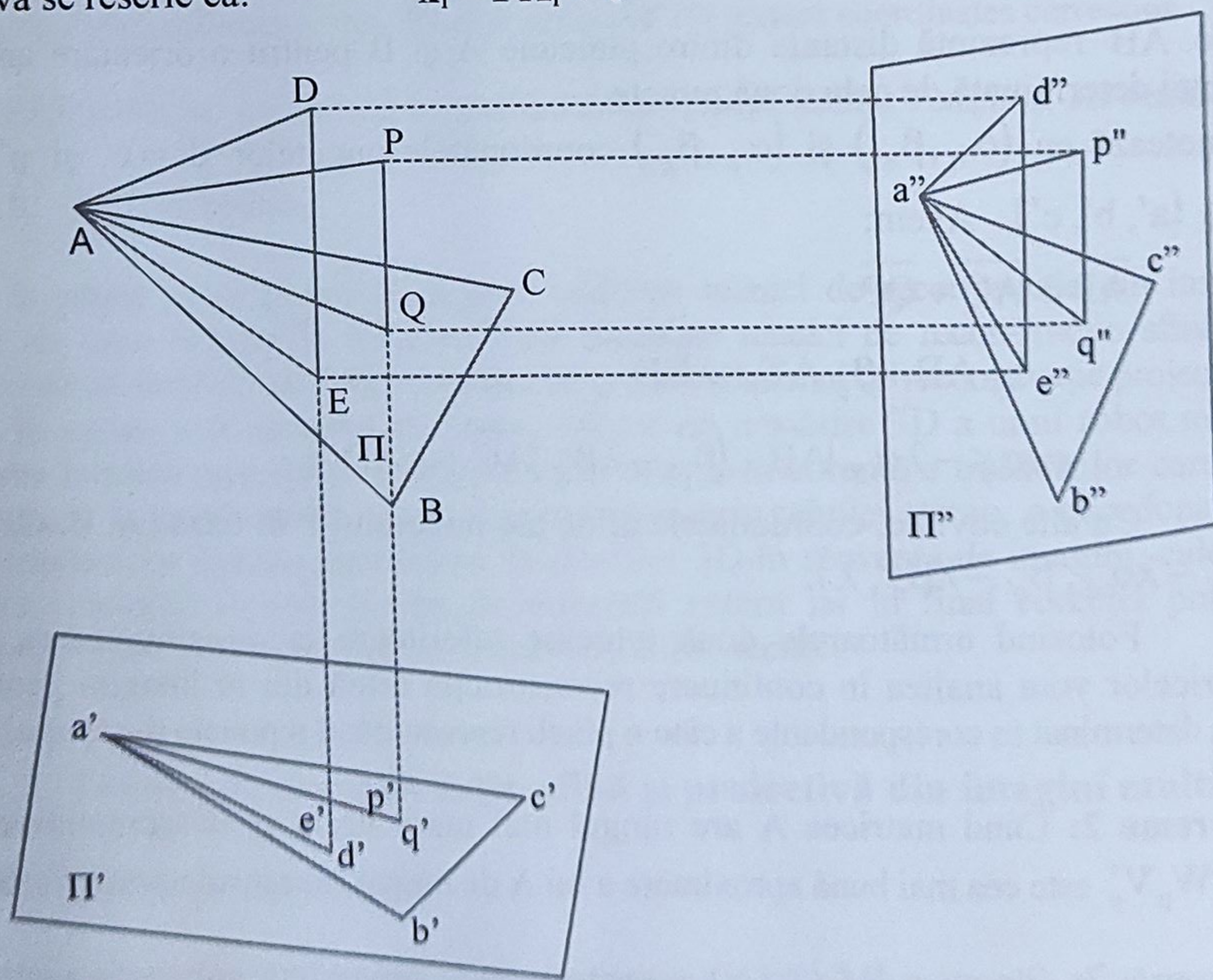


Fig. 1. Construcția geometrică a coordonatelor afine ale punctului P în baza formată de punctele A, B, C, D . (după Forsyth și Ponce³)

³ David A. Forsyth, Jean Ponce, „Computer Vision. A modern approach”, Prentice Hall, 2003.

Algoritmul de reconstrucție afină prezentat de Tomasi⁴ și Hartley, Zisserman⁵, de estimare a m camere afine $\{\mathbf{M}^i, \mathbf{t}^i\}$ și a n puncte tridimensionale $\mathbf{X}_j = (x_j, y_j, z_j)^t$ pornește de la minimizarea distanțelor dintre pixelii estimați $\bar{\mathbf{x}}_j^i = \mathbf{P}_i \mathbf{X}_j + \mathbf{t}^i$ și pixelii corespondenți măsurați \mathbf{x}_j^i :

$$\min_{\mathbf{M}^i, \mathbf{t}^i, \mathbf{X}_j} \sum_{i,j} \|\mathbf{x}_j^i - \bar{\mathbf{x}}_j^i\|^2 = \min_{\mathbf{M}^i, \mathbf{t}^i, \mathbf{X}_j} \sum_{i,j} \|\mathbf{x}_j^i - (\mathbf{P}_i \mathbf{X}_j + \mathbf{t}^i)\|^2 \quad (6)$$

În cele de mai sus $m \geq 2$, $n \geq 4$. Se presupune că pixelii \mathbf{x}_j^i sunt vizibili în toate cele m imagini și că măsurătorile sunt afectate de zgomot gaussian independent de medie nulă.

Minimizarea criteriului (6) în raport cu translațiile \mathbf{t}^i ale celor m sisteme afine conduce la:

$$\frac{\partial}{\partial \mathbf{t}^i} \sum_{k,j} \|\mathbf{x}_j^k - (\mathbf{P}_k \mathbf{X}_j + \mathbf{t}^k)\|^2 = 0$$

adică la sistemul:

$$-\sum_{j=1}^n (\mathbf{x}_j^i - (\mathbf{P}_i \mathbf{X}_j + \mathbf{t}^i)) = n\mathbf{t}^i + \mathbf{P}_i \sum_{j=1}^n \mathbf{X}_j - \sum_{j=1}^n \mathbf{x}_j^i = 0 \quad (7)$$

Prin urmare:

$$\mathbf{t}^i = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j^i - \mathbf{P}_i \frac{1}{n} \sum_{j=1}^n \mathbf{X}_j = \langle \mathbf{x}^i \rangle - \mathbf{P}_i \langle \mathbf{X} \rangle \quad (8)$$

unde s-au folosit notațiile: $\langle \mathbf{x}^i \rangle = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j^i$ și $\langle \mathbf{X} \rangle = \frac{1}{n} \sum_{j=1}^n \mathbf{X}_j$.

Alegând sistemul de coordonate 3D al punctelor $\mathbf{X}_j = (x_j, y_j, z_j)^t$, $j=1, n$, cu originea în punctul $\langle \mathbf{X} \rangle$ relația (8) se reduce la:

$$\mathbf{t}^i = \langle \mathbf{x}^i \rangle \quad (9)$$

Rezultă că în fiecare imagine i , din totalul de m , măsurătorile pixelilor corespondenți se fac într-un sistem având originea în centrul de greutate al celor n pixeli. În criteriul (6) vom avea $\mathbf{t}^i = 0$. Prin urmare, înlocuind în fiecare din seturile \mathbf{x}_j^i valorile măsurate \mathbf{x}_j^i cu $\mathbf{x}_j^i - \langle \mathbf{x}^i \rangle$ criteriul (6) se transformă în:

$$\min_{\mathbf{M}^i, \mathbf{X}_j} \sum_{i,j} \|\mathbf{x}_j^i - \bar{\mathbf{x}}_j^i\|^2 = \min_{\mathbf{M}^i, \mathbf{X}_j} \sum_{i,j} \|\mathbf{x}_j^i - \mathbf{P}_i \mathbf{X}_j\|^2 \quad (10)$$

În forma matriceală relațiile $\bar{\mathbf{x}}_j^i = \mathbf{P}_i \mathbf{X}_j$, $i=1, m$, $j=1, n$ se scriu ca:

⁴ Carlo Tomasi, Takeo Kanade, „Shape and motion from image streams under orthography: A factorization approach”, în *International Journal of Computer Vision*, vol 9, Issue 2, November 1992, pp. 137-154.

⁵ Richard Hartley, Andrew Zisserman, „Multiple View Geometry in Computer Vision”, Cambridge University Press, 2004, Second Edition.

$$\bar{\mathbf{W}} = \begin{bmatrix} \bar{\mathbf{x}}_1^1 & \bar{\mathbf{x}}_2^1 & \cdot & \cdot & \bar{\mathbf{x}}_n^1 \\ \bar{\mathbf{x}}_1^2 & \bar{\mathbf{x}}_2^2 & \cdot & \cdot & \bar{\mathbf{x}}_n^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \bar{\mathbf{x}}_1^m & \bar{\mathbf{x}}_2^m & \cdot & \cdot & \bar{\mathbf{x}}_n^m \end{bmatrix} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \cdot \\ \cdot \\ \mathbf{P}_m \end{bmatrix} [\mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdot \quad \cdot \quad \mathbf{X}_n] \quad (11)$$

Matricea \mathbf{W} a măsurătorilor este:

$$\mathbf{W} = \begin{bmatrix} \mathbf{x}_1^1 & \mathbf{x}_2^1 & \cdot & \cdot & \mathbf{x}_n^1 \\ \mathbf{x}_1^2 & \mathbf{x}_2^2 & \cdot & \cdot & \mathbf{x}_n^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{x}_1^m & \mathbf{x}_2^m & \cdot & \cdot & \mathbf{x}_n^m \end{bmatrix} \quad (12)$$

Deoarece, din cauza erorilor de măsurare, matricea măsurătorilor nu verifică (11), determinarea matricelor \mathbf{P}_i și a punctelor \mathbf{X}_j se realizează prin minimizarea criteriului:

$$E \stackrel{\text{def}}{=} \sum_{i,j} |\mathbf{x}_{ij} - \mathbf{P}_i \mathbf{X}_j|^2 = \sum_j |\mathbf{d}_j - \mathbf{P} \mathbf{X}_j|^2 \quad (13)$$

în raport cu \mathbf{P}_i , $i=1,m$ și vectorii \mathbf{X}_j , $j=1,n$. În relația de mai sus vectorii \mathbf{d}_j sunt coloanele matricei măsurătorilor \mathbf{W} normalizate:

$$\mathbf{W} = \begin{bmatrix} \mathbf{x}_1^1 & \mathbf{x}_2^1 & \cdot & \cdot & \mathbf{x}_n^1 \\ \mathbf{x}_1^2 & \mathbf{x}_2^2 & \cdot & \cdot & \mathbf{x}_n^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{x}_1^m & \mathbf{x}_2^m & \cdot & \cdot & \mathbf{x}_n^m \end{bmatrix} = \begin{bmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \cdot & \cdot & \mathbf{d}_n \\ 1 & 1 & \cdot & \cdot & 1 \end{bmatrix} \quad (14)$$

Minimizăm criteriul E (13) în funcție de \mathbf{X}_j :

$$0 = \frac{\partial E}{\partial \mathbf{X}_j} = \left[2 \frac{\partial}{\partial \mathbf{X}_j} (\mathbf{d}_j - \mathbf{P} \mathbf{X}_j)^t \right] [\mathbf{d}_j - \mathbf{P} \mathbf{X}_j] = -2 \mathbf{P}^t [\mathbf{d}_j - \mathbf{P} \mathbf{X}_j] \quad (15)$$

$$\text{Ceea ce înseamnă că: } \mathbf{X}_j = \mathbf{P}^+ \mathbf{d}_j \quad (16)$$

Matricea \mathbf{P}^+ este pseudo-inversa matricei \mathbf{P} :

$$\mathbf{P}^+ = (\mathbf{P}^t \mathbf{P})^{-1} \mathbf{P}^t \quad (17)$$

Înlocuind \mathbf{X}_j în expresia lui E avem:

$$E = \sum_j |(\mathbf{I} - \mathbf{P} \mathbf{P}^+) \mathbf{d}_j|^2 \quad (18)$$

Se observă că matricea $\mathbf{P} \mathbf{P}^+$ asociată cu un vector \mathbf{d} din \mathbf{R}^{2m} este proiecția ortogonală în subspațiul tridimensional \mathbf{V}_P generat de coloanele lui \mathbf{P} . Rezultă că E

măsoară norma diferenței dintre vectori \mathbf{d}_j și proiecțiile lor ortogonale pe \mathbf{V}_P . Potrivit teoremei 3, E este minimă atunci când \mathbf{V}_P este generat de coloanele matricei \mathbf{U}_3 obținută din cele trei coloane din stânga lui \mathbf{U} , unde $\mathbf{U}\mathbf{W}\mathbf{V}^t$ reprezintă descompunerea în valori singulare ale matricei $\mathbf{D}=[\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n]$. Matricea \mathbf{P} care minimizează criteriul E verifică relația $\mathbf{V}_P = \mathbf{V}_{U_3}$ astfel încât avem: $\mathbf{P} = \mathbf{U}_3$.

Cum $\mathbf{X}_j = \mathbf{P}^+ \mathbf{d}_j$ pentru toate punctele din spațiul afin avem:

$$\mathbf{X} = \mathbf{P}^+ \mathbf{d} = \left[(\mathbf{U}_3^t \mathbf{U}_3)^{-1} \mathbf{U}_3^t \right] [\mathbf{U}\mathbf{W}\mathbf{V}^t] = \mathbf{W}_3 \mathbf{V}_3^t \quad (19)$$

unde, \mathbf{V}_3 este o matrice $3 \times n$ obținută din cele trei coloane din stânga ale matricei

\mathbf{V} iar \mathbf{W}_3 este o matrice 3×3 obținută din valorile singulare corespunzătoare.

În concluzie algoritmul clasic de factorizare pentru calculul reconstrucției afine și a matricelor camerelor este:

1. Se calculează descompunerea SVD a matricei măsurătorilor: $\mathbf{d} = \mathbf{U}\mathbf{W}\mathbf{V}^t$.
2. Se construiesc matricele \mathbf{U}_3 , \mathbf{V}_3 și \mathbf{W}_3 obținute din cele trei coloane din stânga ale matricelor \mathbf{U} , respectiv \mathbf{V} și submatricea corespondentă 3×3 a lui \mathbf{W} .
3. Se calculează: $\mathbf{P} = \mathbf{U}_3$ și $\mathbf{X} = \mathbf{W}_3 \mathbf{V}_3^t$.

\mathbf{P} este o matrice 2×3 care estimează mișcarea camerei iar \mathbf{X} este o matrice $3 \times n$ care estimează structura scenei.

În continuare vom analiza câteva modele de proiecții (camere) afine în vederea trecerii de la reconstrucția afină discutată mai sus la reconstrucția euclidiană.

Presupunem că imaginile sunt preluate de o cameră calibrată astfel încât punctele din imagine sunt reprezentate prin vectorii coordonatelor normalizați. Vom vedea că ecuația proiecției afine $\mathbf{x}_i = \mathbf{P}_i \mathbf{X}$ se aplică și pentru aceste modele. Cu toate acestea, de această dată există unele constrângeri asupra componentelor matricei de proiecție \mathbf{P} .

2.1.1. Proiecția ortografică

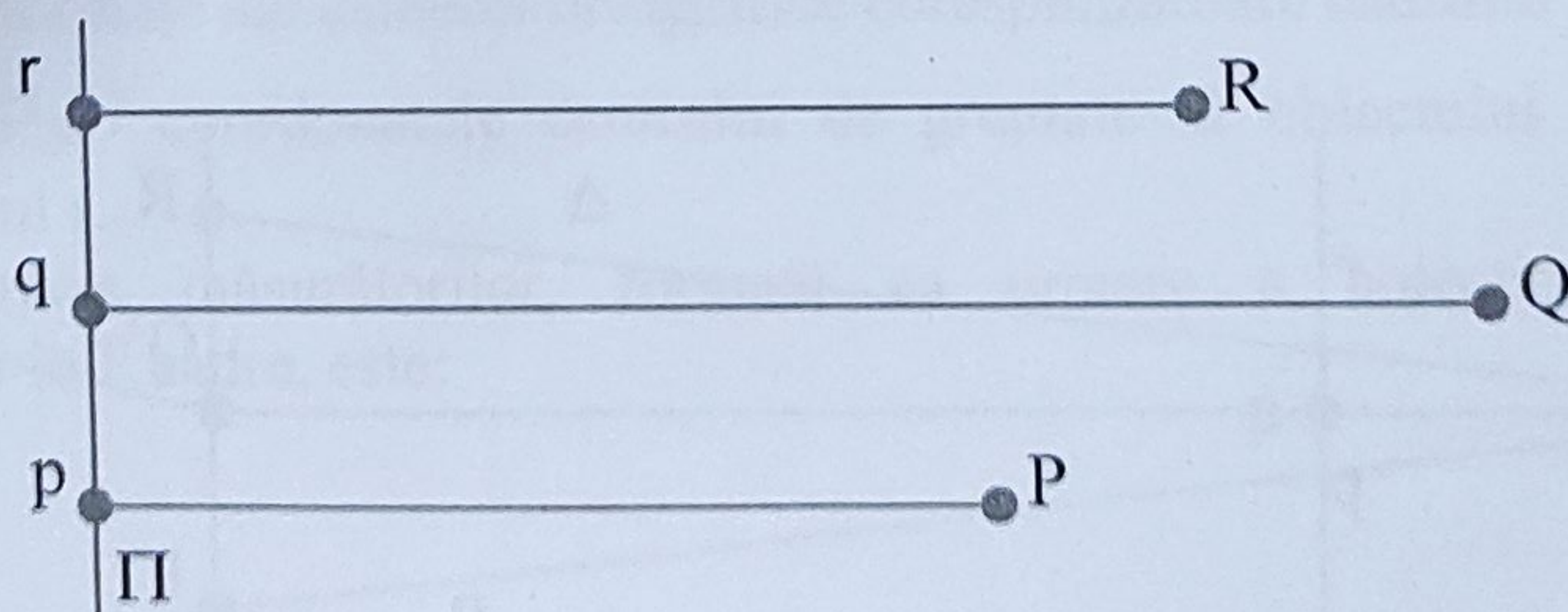


Fig. 2. Proiecția ortografică (după Forsyth și Ponce⁶)

⁶ David A. Forsyth, Jean Ponce, „Computer Vision. A modern approach”, Prentice Hall, 2003.

Pentru o cameră ortografică, matricea P este o parte a matricei de rotație, rândurile \mathbf{a}' și \mathbf{b}' sunt vectori unitate ortogonali. Cu alte cuvinte, o cameră ortografică este o cameră afină cu constrângeri suplimentare:

$$\mathbf{a} \cdot \mathbf{b} = 0 \text{ și } |\mathbf{a}|^2 = |\mathbf{b}|^2 = 1 \quad (20)$$

2.1.2. Proiecția «weak perspective»

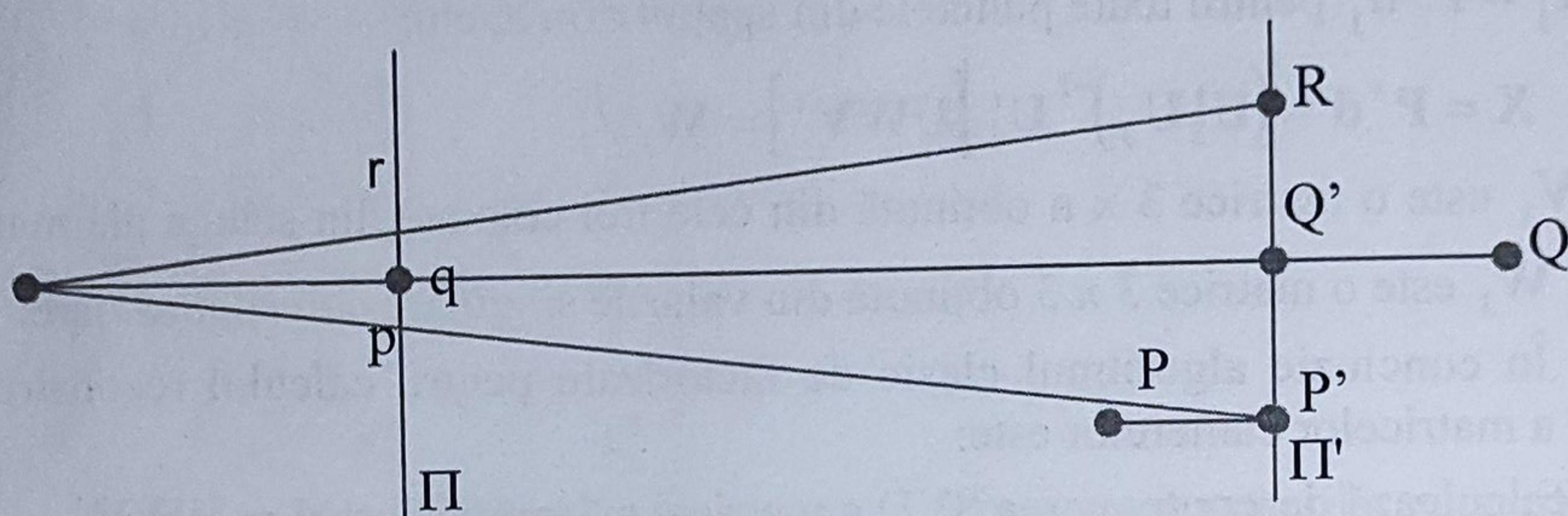


Fig. 3. Proiecția weak perspective (după Forsyth și Ponce⁷)

Proiecția weak perspective este o aproximare a proiecției obișnuite (pinhole camera), construită astfel: se consideră C ca fiind centrul optic al camerei și R un punct de referință al scenei; proiecția weak perspective a unui punct P din scenă este construită în doi pași: prima dată se proiectează ortografic punctul P într-un punct P' din planul Π' paralel cu planul imagine Π , care trece prin punctul R . Proiecția perspectivă este apoi utilizată pentru maparea punctului P' în pixelul p . Deoarece Π' este un plan paralel planului imagine, efectul celui de-al doilea pas este o scalare a coordonatelor imaginii. Rezultă că o cameră weak perspective este o cameră afină îndeplinind condițiile:

$$\mathbf{a} \cdot \mathbf{b} = 0 \text{ și } |\mathbf{a}|^2 = |\mathbf{b}|^2 \quad (21)$$

2.1.3. Proiecția paraperspectivă

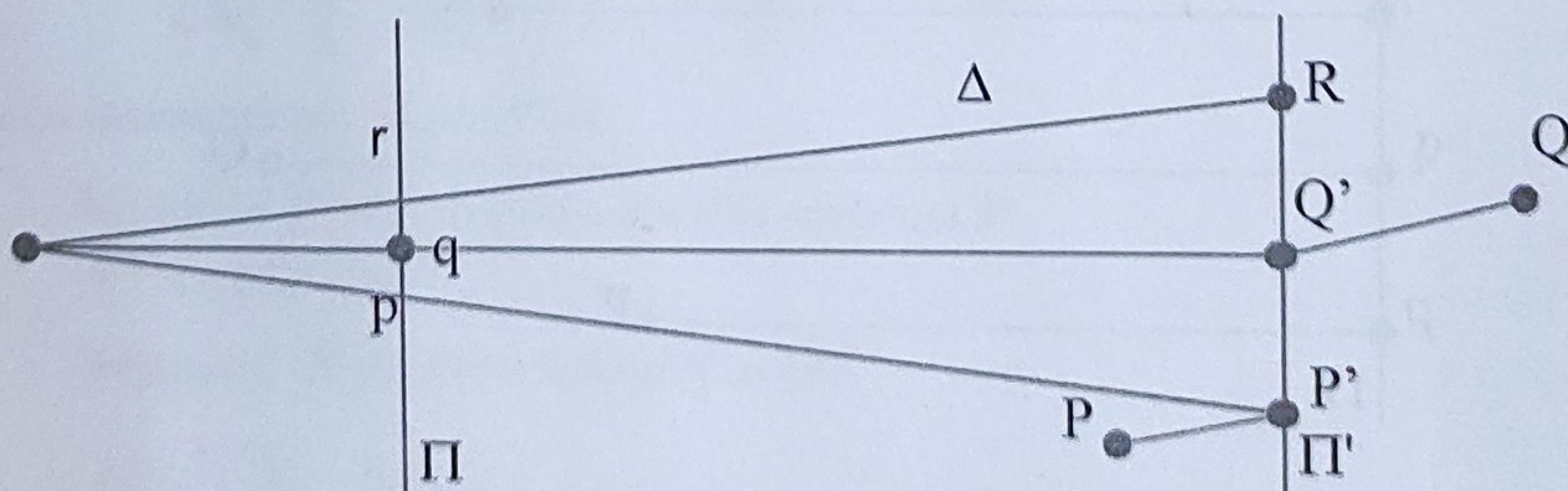


Fig. 4. Proiecția paraperspectivă (după Forsyth și Ponce⁸)

⁷ Idem nota 5.

Proiecția para-perspectivă este o aproximare a camerei pinhole care ia în considerare distorsiunile asociate cu un punct de referință care nu aparține axei optice a camerei. Notând cu Δ dreapta determinată de centrul optic C și punctul de referință R , proiecția paralelă pe direcția lui Δ este prima care se folosește pentru a mapa punctul P' în punctul p din imagine. Se demonstrează că o cameră para-perspectivă este o cameră afină care satisface următoarele restricții:

$$\mathbf{a} \cdot \mathbf{b} = \frac{u_r v_r}{2(1 + u_r^2)} |\mathbf{a}|^2 + \frac{u_r v_r}{2(1 + v_r^2)} |\mathbf{b}|^2 \text{ și } (1 + v_r^2) |\mathbf{a}|^2 = (1 + u_r^2) |\mathbf{b}|^2 \quad (22)$$

unde (u_r, v_r) sunt coordonatele proiecției perspective ale punctului R . Pentru acest model vectorii \mathbf{a} și \mathbf{b} nu formează o bază pentru planul imaginii. În schimb, constituie o bază în planul vectorial ortogonal pe dreapta care unește centrul optic al camerei cu punctul de referință. Cum era de așteptat, constrângerile para-perspective se reduc la constrângerile din cazul anterior (weak perspective) atunci când $u_r = v_r = 0$ iar acestea, la rândul lor, se reduc la constrângerile ortografice atunci când planele Π și Π' coincid.

Să vedem cum se ajunge la o structură euclidiană plecând de la o structură afină realizată prin proiecție ortografică⁹. Presupunem că urmărim un obiect în mișcare cu o cameră care poate fi aproximată afin prin modelul ortografic (obiectul este de dimensiuni mult mai mici în raport cu distanța obiect-cameră). Sunt achiziționate N caracteristici în F cadre, corespondența realizându-se cu unul din algoritmi din literatura de specialitate.

În cele ce urmează folosim notațiile:

- $(u_{fi}, v_{fi})^t$ coordonatele caracteristicii i , $i=1, N$, din cadrul f ;
- $\mathbf{p}_i^t = (X_i, Y_i, Z_i)^t$, coordonatele neomogene ale caracteristicii i în sistemul camerei;
- $\mathbf{s}_i = \begin{bmatrix} \mathbf{p}_i \\ 1 \end{bmatrix}$, coordonatele omogene ale caracteristicii i în sistemul camerei;
- $\mathbf{i}_f = (i_{x_f}, i_{y_f}, i_{z_f})^t$, $\mathbf{j}_f = (j_{x_f}, j_{y_f}, j_{z_f})^t$ primele două rânduri ale matricei ortonormate ale camerei ortografice corespunzătoare cadrului f ;
- $(t_{x_f}, t_{y_f})^t$ coordonatele centrului de greutate al obiectului urmărit pentru cadrul f .

Matricea măsurărilor, formată ca urmare a colecționării celor N caracteristici în F cadre, este:

⁸ Idem nota 5.

⁹ Joao Costeira, Takeo Kanade, „A Multi-body Factorization Method for Motion Analysis” în *International Journal of Computer Vision*, 29(3), 1998, pp. 159-179.

$$\begin{bmatrix} u_{11} & \cdot & \cdot & u_{1N} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ u_{F1} & \cdot & \cdot & u_{FN} \\ v_{11} & \cdot & \cdot & v_{1N} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ v_{F1} & \cdot & \cdot & v_{FN} \end{bmatrix} = \begin{bmatrix} \mathbf{i}_1^t & t_{x1} \\ \cdot & \cdot \\ \cdot & \cdot \\ \mathbf{i}_F^t & t_{xF} \\ \mathbf{j}_1^t & t_{y1} \\ \cdot & \cdot \\ \cdot & \cdot \\ \mathbf{j}_F^t & t_{yF} \end{bmatrix} \begin{bmatrix} s_1 & \cdot & \cdot & s_N \end{bmatrix} \quad (23)$$

În forma compactă (23) este:

$$\mathbf{W} = \mathbf{M}\mathbf{S} \quad (24)$$

unde matricea măsurătorilor a fost notată cu \mathbf{W} , matricele camerelor adică matricea mișcării cu \mathbf{M} iar matricea structurii cu \mathbf{S} .

Calculul matricei de mișcare și a structurii (coordonatele 3D ale caracteristicilor) s-ar rezuma la descompunerea SVD a matricei \mathbf{W} :

$$\mathbf{W} = \mathbf{U} \Sigma \mathbf{V}^t \quad (25)$$

O primă soluție pentru matricea mișcării și cea a structurii ar putea fi:

$$\hat{\mathbf{M}} = \mathbf{U} \Sigma^{\frac{1}{2}}; \quad \hat{\mathbf{S}} = \Sigma^{\frac{1}{2}} \mathbf{V}^t \quad (26)$$

Descompunerea matricei măsurătorilor nu este unică în cazul afin general așa cum s-a arătat. Pentru orice matrice 4x4 nesingulară \mathbf{A} în cazul general este valabilă și soluția: $\mathbf{M} = \hat{\mathbf{M}}\mathbf{A}$; $\mathbf{S} = \mathbf{A}^{-1}\hat{\mathbf{S}}$ întru-cât:

$$\mathbf{M}\mathbf{S} = (\hat{\mathbf{M}}\mathbf{A})(\mathbf{A}^{-1}\hat{\mathbf{S}}) \quad (27)$$

Urmează calculul matricei 4x4 \mathbf{A} :

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_R & \mathbf{a}_t \end{bmatrix} \quad (28)$$

unde primele trei rânduri ale blocului 4x3 \mathbf{A}_R formează o matrice de rotație iar vectorul 4x1 este vectorul translațiilor, care să facă din matricea:

$$\mathbf{M} = \hat{\mathbf{M}}\mathbf{A} = \begin{bmatrix} \hat{\mathbf{M}}\mathbf{A}_R & \hat{\mathbf{M}}\mathbf{a}_t \end{bmatrix} \quad (29)$$

matricea unei camere ortografice.

Condițiile (21) conduc la sistemul:

$$\hat{\mathbf{m}}_i \mathbf{A}_R \mathbf{A}_R^t \hat{\mathbf{m}}_i^t = 1; \quad \hat{\mathbf{m}}_j \mathbf{A}_R \mathbf{A}_R^t \hat{\mathbf{m}}_j^t = 1 \quad (30)$$

$$\hat{\mathbf{m}}_i \mathbf{A}_R \mathbf{A}_R^t \hat{\mathbf{m}}_j^t = 0 \quad (31)$$

necunoscutele fiind elementele matricei \mathbf{A}_R .

În ecuațiile de mai sus $\hat{\mathbf{m}}_i$, $i=1,F$; $\hat{\mathbf{m}}_j$, $j=F+1,2F$, sunt rândurile matricei $\hat{\mathbf{M}}$.

Vectorul translațiilor \mathbf{a}_t se calculează în felul următor. Cum proiecția centrului de greutate al obiectului este totuna cu centrul de greutate al proiecțiilor caracteristicilor avem:

$$\bar{\mathbf{w}} = \begin{bmatrix} \frac{1}{N} \sum u_{li} \\ \vdots \\ \frac{1}{N} \sum v_{Fi} \end{bmatrix} = \mathbf{M}\bar{\mathbf{s}} \quad (32)$$

$$= \begin{bmatrix} \hat{\mathbf{M}}\mathbf{A}_R & \hat{\mathbf{M}}\mathbf{a}_t \end{bmatrix} \begin{bmatrix} \bar{\mathbf{p}} \\ 1 \end{bmatrix} \quad (33)$$

unde $\bar{\mathbf{p}} = \frac{1}{N} \sum \mathbf{p}_i$ este centrul de greutate al caracteristicilor 3D ale obiectului.

Cum originea sistemului de coordonate atașat obiectului poate fi aleasă în centrul de greutate al caracteristicilor acestuia avem $\bar{\mathbf{p}} = \mathbf{0}$, ceea ce înseamnă că \mathbf{a}_t poate fi calculat din:

$$\bar{\mathbf{w}} = \hat{\mathbf{M}}\mathbf{a}_t \quad (34)$$

Soluția este:

$$\mathbf{a}_t = (\hat{\mathbf{M}}^t \hat{\mathbf{M}})^{-1} \hat{\mathbf{M}}^t \bar{\mathbf{w}} = \quad (35)$$

$$= \sum^{-1/2} \mathbf{U}^t \bar{\mathbf{w}} \quad (36)$$

Celelalte două cazuri de proiecție afină sunt tratate în aceeași manieră, cu excepția faptului că în cazul proiecției weak perspective condițiile $\mathbf{a} \cdot \mathbf{b} = 0$ și $|\mathbf{a}|^2 = |\mathbf{b}|^2$ pentru $m-1$ imagini vor înlocui cele $3m-3$ condiții (21) iar în cazul proiecției paraperspective se vor utiliza condițiile (22) în locul celor $3m-3$ condiții (20).

Am presupus până acum că toate cele n caracteristici punctuale din cele m imagini aparțin aceluiași obiect, deci se deplasează cu aceeași viteză. Pentru cazul în care caracteristicile aparțin unor k obiecte având mișcări diferite sunt prezentate două metode de segmentare a punctelor în mulțimi de puncte reprezentând același obiect. Problema se reduce la factorizarea mai multor matrice ale măsurătorilor, compuse din corespondențele caracteristicilor aceluiași obiect.

a) Forma redusă a matricei de date

Definim matricea măsurătorilor ca fiind:

$$\mathbf{D} = \begin{bmatrix} p_{11} & \dots & p_{1n} \\ \dots & \dots & \dots \\ p_{m1} & \dots & p_{mn} \\ 1 & \dots & 1 \end{bmatrix} \quad (37)$$

De data aceasta, \mathbf{D} nu are rangul 4. În schimb, sub-matricile formate din coloanele corespunzătoare fiecărui obiect vor avea cel mult rangul 4. Rangul maximal al matricei globale de date va fi $4k$. Cu alte cuvinte, coloanele lui \mathbf{D} corespunzătoare

pentru fiecare obiect se află în subspații cel mult patru-dimensionale. Construirea, formei reduse a lui D va identifica aceste subspații și vectori coloană care aparțin aceluiași obiect. Gear oferă¹⁰ mai multe metode pentru calcularea formei reduse folosind metoda de eliminare Gauss-Jordan și de reducere QR.

b) Matricea formei de interacțiune

Abordarea prezentată în secțiunea anterioară se bazează pe structura afină a imaginii afine. Costeira și Kanade au propus¹¹ o metodă diferită de factorizare a matricei de date.

Presupunând că se urmărește deplasarea a două obiecte în mișcare și că s-a realizat segmentarea caracteristicilor astfel încât se cunoaște apartenența lor la obiecte vom avea o matrice a măsurătorilor în forma canonică:

$$\mathbf{W}^* = [\mathbf{W}_1 | \mathbf{W}_2] \quad (38)$$

unde submatricea \mathbf{W}_1 are N_1 coloane corespunzătoare celor N_1 caracteristici ale primului obiect iar submatricea \mathbf{W}_2 are N_2 coloane corespunzătoare celor N_2 caracteristici ale celui de al doilea obiect.

Fiecare submatrice a măsurătorilor poate fi factorizată ca în cazul urmăririi mono-obiect, relația (25):

$$\mathbf{W}_1 = \mathbf{U}_1 \sum_1 \mathbf{V}_1^t = \quad (39)$$

$$= \mathbf{M}_1 \mathbf{S}_1 = (\hat{\mathbf{M}}_1 \mathbf{A}_1)(\mathbf{A}_1^{-1} \hat{\mathbf{S}}_1) \quad (40)$$

unde $l=1,2$.

Ecuția (38) are așadar următoarea formă:

$$\mathbf{W}^* = [\mathbf{M}_1 | \mathbf{M}_2] \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{bmatrix} = \quad (41)$$

$$= [\hat{\mathbf{M}}_1 | \hat{\mathbf{M}}_2] \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \mathbf{A}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{S}}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{S}}_2 \end{bmatrix} \quad (42)$$

Folosind notațiile:

$$\hat{\mathbf{W}}^* = [\hat{\mathbf{M}}_1 | \hat{\mathbf{M}}_2] \quad (43)$$

$$\hat{\mathbf{S}}^* = \begin{bmatrix} \hat{\mathbf{S}}_1 & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{S}}_2 \end{bmatrix} \quad (44)$$

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix} \quad (45)$$

$$\mathbf{U}^* = [\mathbf{U}_1 | \mathbf{U}_2] \quad (46)$$

¹⁰ C. William Gear, „Feature grouping in moving objects” în *Proceedings of the 1994 IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, 1994., pp. 214-219.

¹¹ Joao Costeira, Takeo Kanade, „A Multi-body Factorization Method for Motion Analysis” în *International Journal of Computer Vision*, 29(3), 1998, pp. 159-179.

$$\Sigma^* = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \quad (47)$$

$$V^{*t} = \begin{bmatrix} V_1^t & 0 \\ 0 & V_2^t \end{bmatrix} \quad (48)$$

Se obține o factorizare similară celei urmăririi unui singur obiect:

$$W^* = M^* S^* \quad (49)$$

$$S^* = A^{*-1} \hat{S}^* = A^{*-1} \Sigma^{*\frac{1}{2}} V^{*t} \quad (50)$$

$$M^* = \hat{M}^* A^* = U^* \Sigma^{*\frac{1}{2}} A^* \quad (51)$$

Din relația (38) rezultă ca rangul matricei măsurătorilor pentru două obiecte diferite are rangul cel mult 8. În continuarea paragrafului vom considera ca cele două obiecte sunt tridimensionale prin urmare matricea W^* are rangul 8, întrucât submatricele corespunzătoare celor două obiecte au rangul 4. În general, rangul unei submatrice a măsurătorilor reprezentând un plan în mișcare este 3 iar rangul unei submatrice rezultate din urmărirea unei drepte în spațiu e 2.

În realitate nu se cunoaște apartenența caracteristicilor la obiecte, prin urmare coloanele matricei măsurătorilor reprezentând un obiect pot fi intercalate între coloanele reprezentând caracteristicile celui alt obiect. Totuși matricea măsurătorilor poate fi descompusă SVD în forma cunoscută:

$$W = U \Sigma V^t \quad (52)$$

urmând a fi determinată matricea A în așa fel încât matricea formelor să aibă expresia (50) iar cea a mișcării forma (51). Totuși din cauză că acest lucru presupune segmentarea caracteristicilor deja efectuată, trebuie aplicată mai întâi o operație de regroupare a coloanelor matricei W inițiale în așa fel încât, în ipoteza mișcării a două obiecte 3D, aceasta să fie defalcată în două submatrice de rang 4. O idee ar fi ca începând cu primele patru coloane liniar independente ale matricei W să fie depistate și grupate adiacent toate celelalte coloane liniar dependente de baza celor patru. Procedul continuă recursiv până se obține în final o segmentare a matricei inițiale în două submatrice de rang patru, în exemplul nostru. Totuși procedura, atunci când nu este cunoscut numărul obiectelor și tipul acestora (obiect 3D, plan sau dreaptă), poate deveni foarte complexă. Costeira și Kanade au propus¹² o metodă de grupare pe obiecte a coloanelor matricei W care pornește de la informațiile conținute într-o matrice Q a interacțiunii formelor definită ca fiind:

$$Q = VV^t \quad (53)$$

unde V este matricea rezultată din descompunerea SVD (52) a matricei inițiale a măsurătorilor. În cazul particular al deplasării celor două obiecte 3D matricea Q are dimensiunea $(N1+N2) \times (N1+N2)$.

¹² Idem nota 10.

Matricea Q are mai multe proprietăți interesante:

1. **Unicitatea.** Se obține unic, prin descompunerea SVD (52) a matricei măsurătorilor inițiale, descompunere care dă matricea V .

2. **Fiecare element al lui Q conține informații referitoare la faptul că o anumită pereche de caracteristici aparține sau nu aceluiași obiect.** Din moment ce forma canonică W^* a matricei măsurătorilor, adică forma în care coloanele reprezentând același obiect sunt grupate adiacent, se obține prin permutarea coloanelor matricei W inițiale și matricea V^{*t} rezultă din permutarea aceluiași coloane ale matricei V^t din descompunerea (52) a matricei inițiale W . De asemenea și forma canonică $Q^* = V^* V^{*t}$ a matricei interacțiunii formelor va rezulta din permutări corespunzătoare de rânduri și coloane. Introducând (50) în $Q^* = V^* V^{*t}$ se obține:

$$Q^* = V^* V^{*t} \quad (54)$$

$$= S^{*t} A^{*t} \Sigma^{*-1} A^* S^* \quad (55)$$

$$= S^{*t} (A^{*-1} \Sigma^* A^{*-t})^{-1} S^* \quad (56)$$

$$= S^{*t} [(A^{*-1} \Sigma^{*-1/2} V^{*t})(V^* \Sigma^{*-1/2} A^{*-t})]^{-1} S^* \quad (57)$$

$$= S^{*t} (S^* S^{*t})^{-1} S^* \quad (58)$$

$$= \begin{bmatrix} S_1^t & 0 \\ 0 & S_2^t \end{bmatrix} \begin{bmatrix} \Lambda_1^{-1} & 0 \\ 0 & \Lambda_2^{-1} \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \quad (59)$$

unde Λ_1 și Λ_2 sunt matrice 4x4 reprezentând momentele de inerție ale fiecărui obiect.

Rezultă că forma canonică Q^* a matricei interacțiunii formelor Q are o structură bloc diagonală în care fiecare element Q_{ij}^* poate avea următoarele valori:

$$Q_{ij}^* = \begin{cases} s_{1i}^t \Lambda_1^{-1} s_{1j} \\ s_{2i}^t \Lambda_2^{-1} s_{2j} \\ 0 \end{cases} \quad (60)$$

Dacă $Q_{ij}^* = s_{1i}^t \Lambda_1^{-1} s_{1j}$ trăsăturile i și j aparțin obiectului 1, dacă $Q_{ij}^* = s_{2i}^t \Lambda_2^{-1} s_{2j}$ trăsăturile i și j aparțin obiectului 2 iar dacă $Q_{ij}^* = 0$ trăsăturile i și j aparțin unor obiecte diferite.

3. **Forma canonică este invariabilă în raport cu numărul obiectelor.** În cele de mai sus matricea interacțiunii formelor a fost construită pentru cazul urmăririi a două obiecte. În cazul general, pentru M obiecte, în formă canonică matricea va avea M blocuri diagonale:

$$Q^* = \begin{bmatrix} S_1^t \Lambda_1^{-1} S_1 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & S_k^t \Lambda_k^{-1} S_k & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & S_M^t \Lambda_M^{-1} S_M \end{bmatrix} \quad (61)$$

Ca și mai sus, atunci când caracteristicile i și j nu aparțin aceluiași obiect elementul Q_{ij}^* este nul. Proprietatea rămâne valabilă și pentru elementele matricei Q calculate din matricea V a descompunerii SVD a matricei măsurătorilor W în forma inițială, în care coloanele caracteristicilor nu sunt ordonate după apartenența la obiecte.

4. O altă proprietate importantă a matricei Q constă în **independența în raport cu tipul mișcării fiecărui corp**, din cauză că elementele matricei conțin informații exclusiv asupra structurii obiectelor (elementele matricei V).

5. **Invarianța la scalarea imaginilor**. Atunci când coordonatele caracteristicilor din imagini sunt scalate matricea măsurătorilor (23) are descompunerea:

$$\begin{bmatrix} u_{11} & \cdot & \cdot & u_{1N} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ u_{F1} & \cdot & \cdot & u_{FN} \\ v_{11} & \cdot & \cdot & v_{1N} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ v_{F1} & \cdot & \cdot & v_{FN} \end{bmatrix} = \begin{bmatrix} c_1 \mathbf{i}_1^t & c_1 t_{x1} \\ \cdot & \cdot \\ \cdot & \cdot \\ c_F \mathbf{i}_F^t & c_F t_{xF} \\ c_1 \mathbf{j}_1^t & c_1 t_{y1} \\ \cdot & \cdot \\ \cdot & \cdot \\ c_F \mathbf{j}_F^t & c_F t_{yF} \end{bmatrix} \begin{bmatrix} s_1 & \cdot & \cdot & s_N \end{bmatrix} \quad (62)$$

unde coeficienții de scalare sunt elementele matricei diagonale:

$$C_{2F \times 2F} = \text{diag}(c_1, \dots, c_F, c_1, \dots, c_F)$$

Invarianța la scalare rezultă din rescrierea relației (62) în forma:

$$\tilde{W} = CMS = [C_1 M_1 | C_2 M_2] \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \quad (63)$$

6. **Invarianța la sistemele de coordonate ale imaginilor** Să presupunem că matricei formelor S a obiectului k i se aplică transformarea 4×4 T :

$$S' = TS \quad (64)$$

Sub-matricea bloc-diagonală corespunzătoare obiectului k în matricea interacțiunii formelor va fi deci:

$$S'^t(S'S'^t)^{-1}S' = (TS)^t[(TS)(TS)^t]^{-1}(TS) = S'(SS'^t)^{-1}S \quad (65)$$

7. Invarianța la tipul obiectelor urmărite Forma bloc diagonală a matricei interacțiunii formelor este invariantă la tipul obiectelor urmărite. Rangul matricei formelor S este 2 pentru o dreaptă, 3 pentru un plan sau 4 pentru un obiect 3D. Cum Q este invariantă la sistemul de coordonate al imaginii, dacă obiectul k este o dreaptă S_k va fi o matrice $2 \times N_k$ iar Λ_k va fi o matrice 2×2 . Similar, dacă obiectul este un plan S_k va fi o matrice $3 \times N_k$ iar Λ_k va fi o matrice 3×3 . În cazul general, pentru un obiect 3D, S_k este o matrice $4 \times N_k$ iar Λ_k o matrice 4×4 . Rangul matricei Q depinde de tipul obiectelor urmărite nu însă și forma bloc diagonală a matricei Q^* .

Așa cum este ilustrat mai jos, continuând exemplul celor două obiecte, reordonarea coloanelor matricei măsurătorilor W inițiale în funcție de apartenența la obiecte este echivalentă cu bloc-diagonalizarea matricei interacțiunii formelor Q .

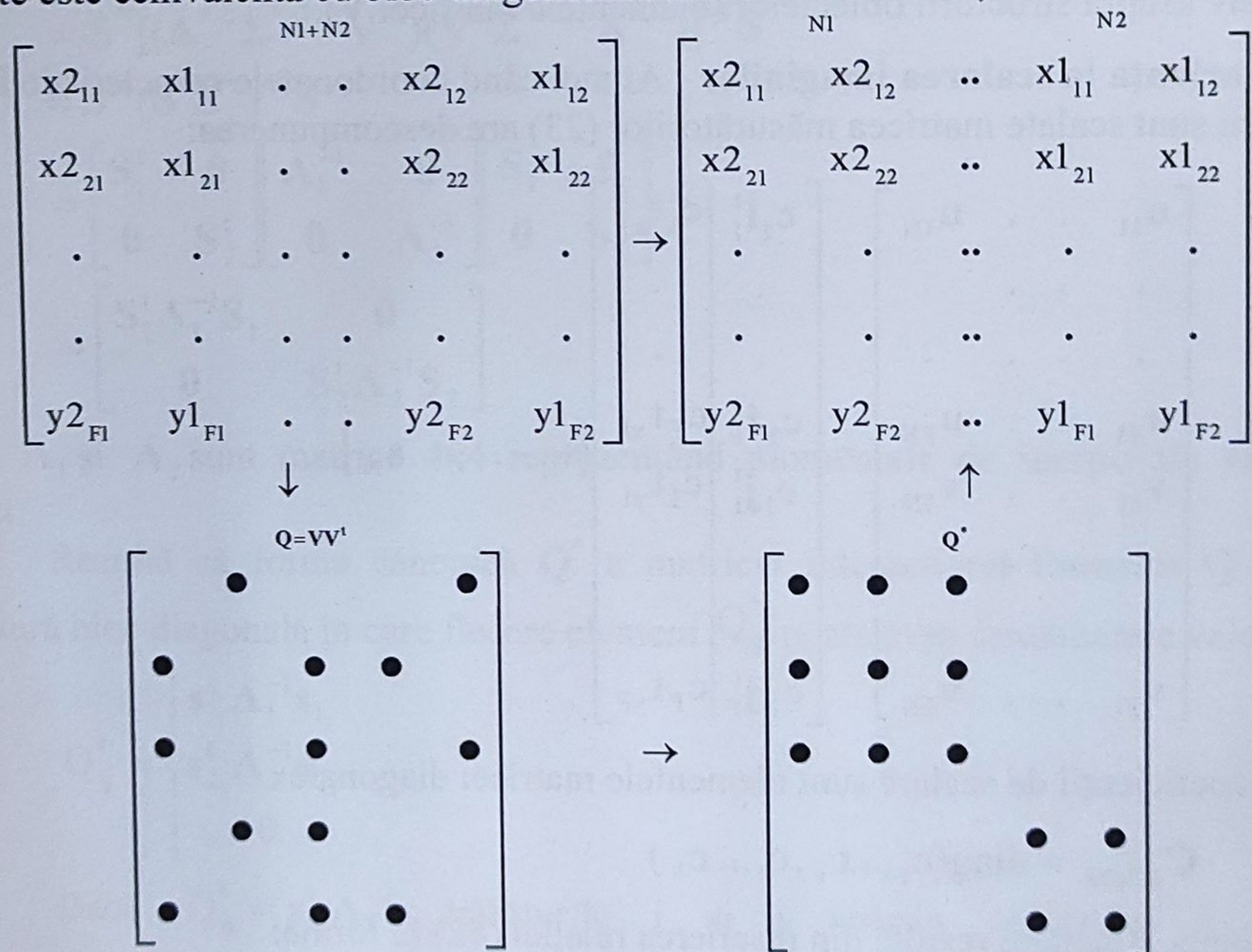


Fig. 5. Segmentarea caracteristicilor (după Costeira și Kanade¹³)

În lucrarea „A Multi-body Factorization Method for Motion Analysis”¹⁴ este propus un algoritm de bloc-diagonalizare a matricei interacțiunii formelor care mai întâi realizează sortarea coloanelor și rândurilor pentru crearea unor blocuri diagonale inițiale iar apoi segmentează sub-blocurile obținute în primul pas în sub-matrice corespunzătoare fiecărui obiect.

¹³ Idem nota 10.

¹⁴ Idem nota 8

În primul pas, diagonalizarea se realizează iterativ alegându-se acele permutări care minimizează energia posibilelor blocuri ne-diagonale ale matricei Q . Presupunem că la pasul k , prin rearanjarea a k coloane și rânduri ale matricei Q inițiale, dispunem de o sub-matrice $k \times k$ Q^{*k} . Crearea sub-matricei $(k+1) \times (k+1)$ Q^{*k+1} se realizează prin permutări de coloane și rânduri în afara primelor k în așa fel încât să fie maximizat criteriul:

$$C_j^k = \sum_{i=1}^k Q_{ij}^2 \quad (66)$$

unde $j=k+1, k+2, \dots, N$.

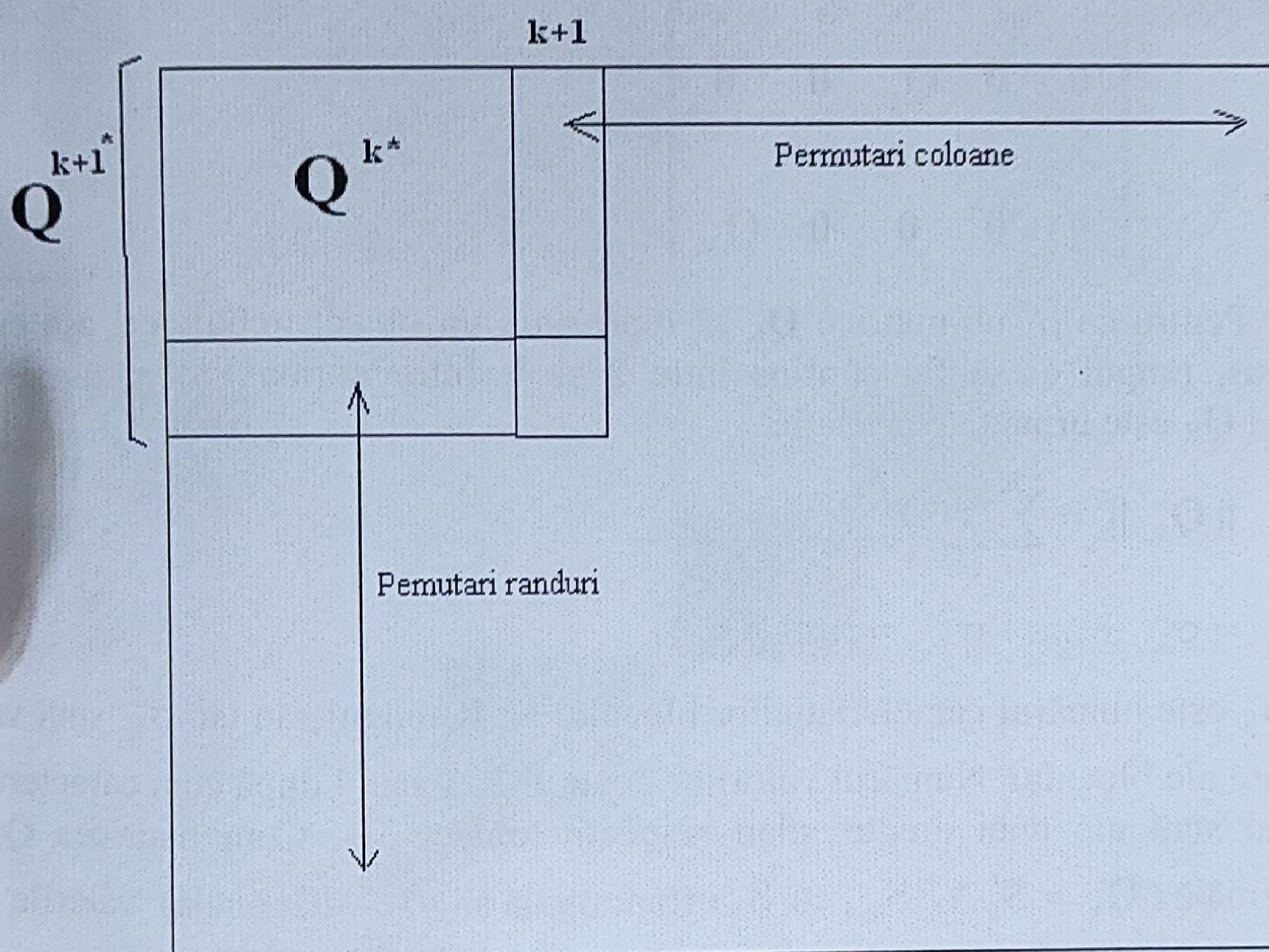


Fig. 6. Algoritmul de sortare. În iterația k coloanele $k+1, \dots, N$ sunt permutate astfel încât coloana care maximizează criteriul 1.66 să completeze, împreună cu rândul corespunzător, sub-matricea Q^{*k} pentru formarea sub-matricei Q^{*k+1} (după Costeira și Kanade¹⁵).

Sub-matricea Q^{*k+1} se formează prin adăugarea în poziția $k+1$, a coloanei, dintre coloanele $k+1, \dots, N$, care maximizează criteriul (66). De asemenea este adăugat și rândul cu același indice rezultat din permutarea de rânduri.

În pasul al doilea al metodei Costeira-Kanade matricea obținută în primul pas, având forma (67), va fi segmentată în sensul reîmpărțirii sub-blocurilor diagonale obținute în primul pas și alocarea fiecărui sub-bloc la un obiect. Prin urmare, în final se obține o matrice Q formată din atâtea blocuri diagonale câte obiecte sunt, fiecare sub-bloc matriceal având rangul cel mult 4, mai exact cuprins între 2 și 4.

În urma sortării din primul pas se obține următoarea matrice a interacțiunii formelor:

¹⁵ Joao Costeira, Takeo Kanade, „A Multi-body Factorization Method for Motion Analysis” în *International Journal of Computer Vision*, 29(3), 1998, pp. 159-179.

$$Q^* = \begin{bmatrix} S_1^t \Lambda_1^{-1} S_1 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & S_k^t \Lambda_k^{-1} S_k & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & S_M^t \Lambda_M^{-1} S_M \end{bmatrix} \quad (67)$$

$$= \begin{bmatrix} Q_1 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & Q_k & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & Q_M \end{bmatrix} \quad (68)$$

Pentru ca o sub-matrice Q_K să reprezinte un obiect trebuie ca, așa cum s-a mai spus, rangul ei să fie cuprins între 2 și 4. Între norma Frobenius și rangul matricei Q_K este următoarea relație:

$$\|Q_K\|^2 = \sum_{i=1}^{N_K} \sum_{j=1}^{N_K} Q_{ij}^2 \quad (69)$$

$$= \sigma_{K_1}^2 + \dots + \sigma_{K_R}^2 = \text{rang}(Q_K) \quad (70)$$

unde N_K este numărul caracteristicilor blocului K , R rangul său iar $\sigma_{K_i}^2$ sunt valorile singulare ale blocului. Numărul valorilor poate fi 2, 3 sau 4 după cum caracteristicile blocului sunt ale unei drepte, plan respectiv obiect 3D. Cum matricea Q_K este ortonormală ($Q_k = S_k^t \Lambda_k^{-1} S_k$, cu S_k ortonormală și Λ_k^{-1} diagonală) valorile singulare ale acesteia $\sigma_{K_i}^2$ sunt unitare prin urmare pătratul normei Frobenius al matricei Q_K este totuna cu rangul ei.

În loc să calculăm (69) considerăm suma:

$$\varepsilon(m) = \sum_{i=1}^m \sum_{j=1}^m Q_{ij}^2 \quad (71)$$

Coloana care va mări cu o unitate suma $\varepsilon(m)$ este un potențial delimitator de obiect, în situația în care suma precedentă este cel puțin 2.

Segmentarea în acest fel însă nu conduce la o soluție unică. De pildă, pentru o matrice Q^* de rang 8, generată de două obiecte 3D, avem posibilitățile figurate mai jos:

- două sub-matrice de rang 4, reprezentând fiecare un obiect 3D – fig.7.a;
- o sub-matrice de rang 3 asociată unui plan și două sub-matrice de rang 2 asociate fiecare la câte o dreaptă – fig.7.b.;
- patru sub-matrice de rang 2, reprezentând fiecare câte o dreaptă-fig.7.c.;

- o sub-matrice de rang 4 asociată unui obiect 3D și două sub-matrice de rang 2 asociate unor drepte-fig.7.d.

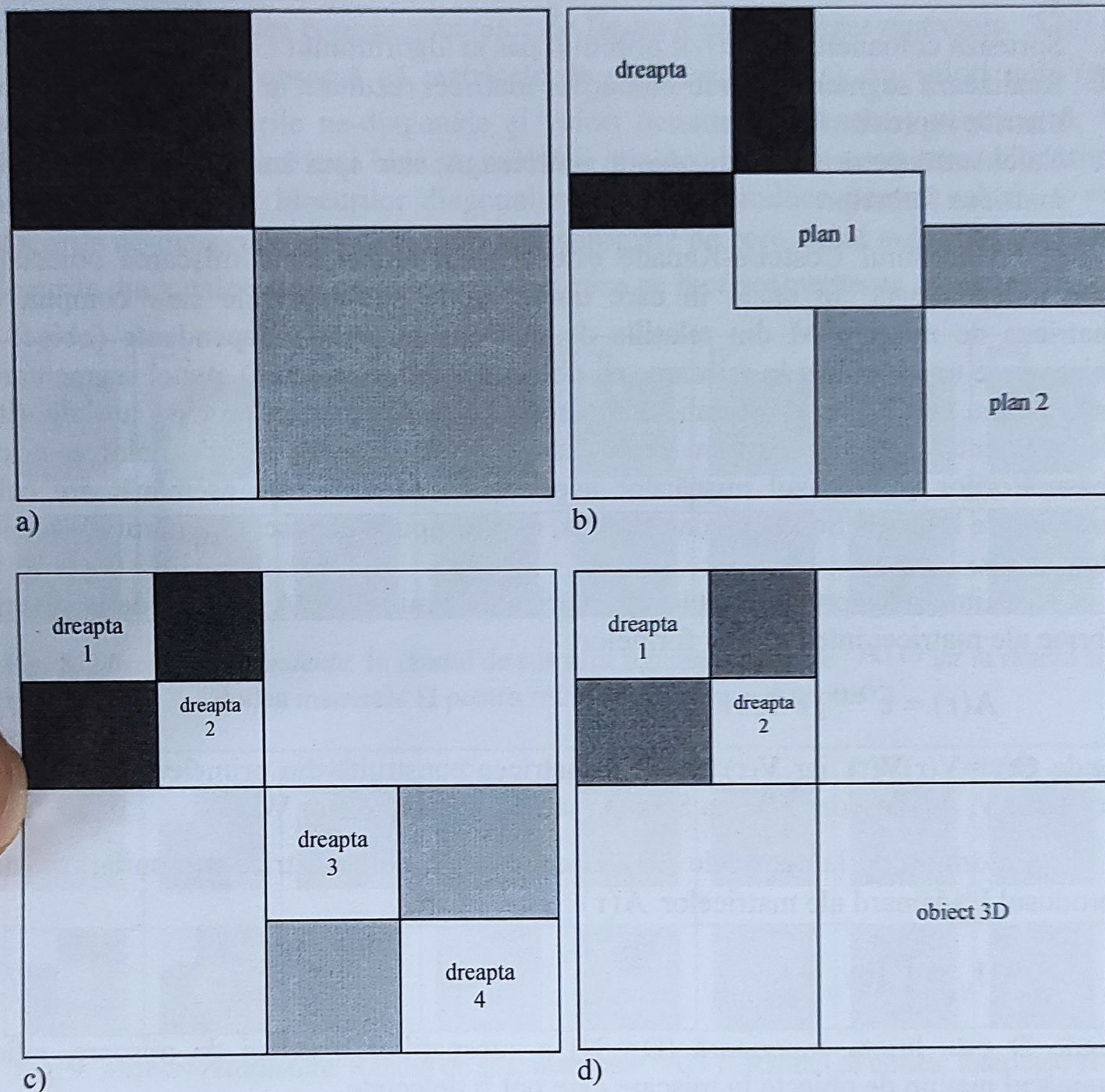


Fig. 7. O matrice Q^* de rang 8, generată în realitate de două obiecte 3D, are patru posibile segmentări în sub-matrice asociate obiectelor (după Costeira și Kanade¹⁶)

Întrucât energia totală a matricei Q^* este:

$$\epsilon(N) = \|Q^*\|_F^2 = \sum_{i=1}^N \sum_{j=1}^N Q_{ij}^2 = \text{rang}(Q) \quad (72)$$

se va alege soluția care concentrează cât mai mult din energia (72) în sub-blocuri, deci minimizează energia blocurilor ne-diagonale.

În concluzie algoritmul de reconstrucție afină în cazul multi-obiect este următorul:

1. Construcția matricei măsurătorilor W , urmarea unui proces de tracking.
2. Calculează $r = \text{rang}(W)$.

¹⁶ Idem nota 14.

3. Descompunerea SVD a matricei măsurătorilor: $W = U \Sigma V^t$.
4. Construiește matricea interacțiunii formelor Q din primele r rânduri ale matricei V^t .
5. Sortează coloanele Q potrivit primului pas al algoritmului Costeira-Kanade.
6. Realizează segmentarea sub-blocurilor matricei rezultate în pasul anterior în sub-matrice reprezentând obiecte.
7. Calculează pentru fiecare obiect matricea A_i , iar apoi matricea de mișcare și matricea formelor.

Algoritmul Costeira-Kanade este eficient atunci când mișcarea obiectelor este independentă. În cazul în care unele dintre sub-matricele care compun de matricea de mișcare M din relațiile de mai sus nu sunt independente (obiect în mișcare pe un alt obiect în mișcare sau obiect nerigid în mișcare) atunci segmentarea propusă nu mai dă rezultate satisfăcătoare. Kim și Agapito au propus¹⁷ un algoritm de segmentare a mișcării care realizează o bună separare a coloanelor matricei măsurătorilor și în cazul mișcărilor dependente. Metoda este asemănătoare celei propuse de Lauer și Schnorr¹⁸ dar, în plus, nu presupune cunoscut numărul obiectelor de urmărit.

Kim și Agapito construiesc¹⁹ matrice de similaritate A pornind de la diverse forme ale matricei interacțiunii formelor:

$$A(r) = e^{Q(r)}, r=2, N \quad (73)$$

unde $Q(r)=V(r)V(r)^t$ iar $V(r)$ este sub-matricea construită din primele r coloane ale matricei V , din descompunerea SVD a matricei măsurătorilor W .

Pornind de la matricele (73) se construiesc noile matrice de similaritate din produsul Hadamard ale matricelor $A(r)$:

$$H_{ij} = \prod_{r=2}^D A_{ij}(r) \quad (74)$$

unde D este limita superioară ($D \leq N$) a subspațiilor matricei de mișcare, adică numărul maxim de obiecte în mișcare care pot fi detectate.

Din ultimile două ecuații rezultă:

$$H_{ij} = \exp\left(\sum_{r=2}^D V_i(r)V_j(r)^t\right) = \exp\left(\sum_{r=2}^D \cos\theta(r)_{ij}\right) \leq e^{D-1} \quad (75)$$

unde $V_i(r)$ este coloana i a matricei $V(r)$ iar $\theta(r)_{ij}$ este unghiul dintre vectorii $V_i(r)$ și $V_j(r)$.

¹⁷ Jae-Hak Kim, Lourdes Agapito, „Motion Segmentation using the Hadamard Product and Spectral Clustering” în *Proceedings of the 2009 international conference on Motion and video computing (WMVC 2009)*, December 2009, Snowbird, Utah, USA, pp. 126-133.

¹⁸ Fabien Lauer, Christoph Schnorr, „Spectral clustering of linear subspaces for motion segmentation” în *The 12th IEEE International Conference on Computer Vision (ICCV'09)*, 2009, pp. 678-675.

¹⁹ Idem nota 16.

Kim și Agapito au prezentat²⁰ rezultatele analizei mișcării a trei obiecte, urmărite de o cameră afină statică. Mai întâi obiectele sunt deplasate utilizând rotații și translații independente iar din caracteristicile corespondente din lanțul de imagini este construită matricea corespondențelor. În figura 8 sunt ilustrate matricele $A(r)$ și H pentru $r=2,9$. Se observă că matricele de similaritate $A(r)$ au valori nule (gri închis) pentru blocurile ne-diagonale și valori nenule (gri deschis) pentru $r=8, 9$. Matricele H se pretează mai bine segmentării mișcării întrucât conturarea blocurilor nule ne-diagonale și blocurilor diagonal ne-nule se produce pentru valori $r \geq 5$. Blocurile ne-diagonale au valori mult mai apropiate de zero decât matricele $A(r)$ iar blocurile diagonale valori ne-nule mai mari ceea ce face segmentarea mai simplă.

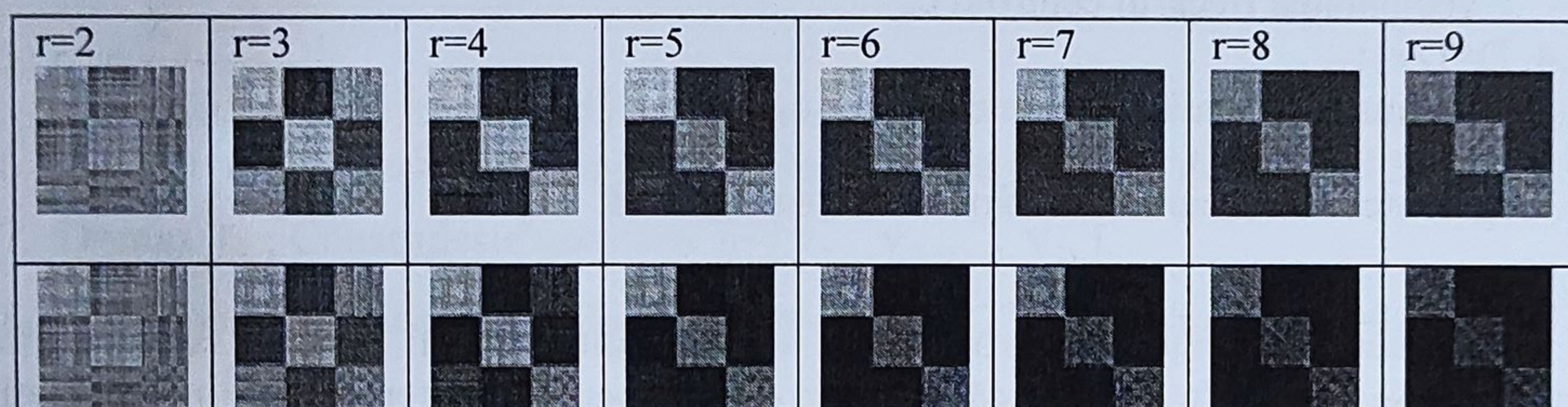


Fig. 8. Mișcări independente. În rândul de sus sunt figurate matricele $A(r)$ iar în rândul al doilea matricele H pentru $r=2,9$ (după Kim și Agapito²¹)

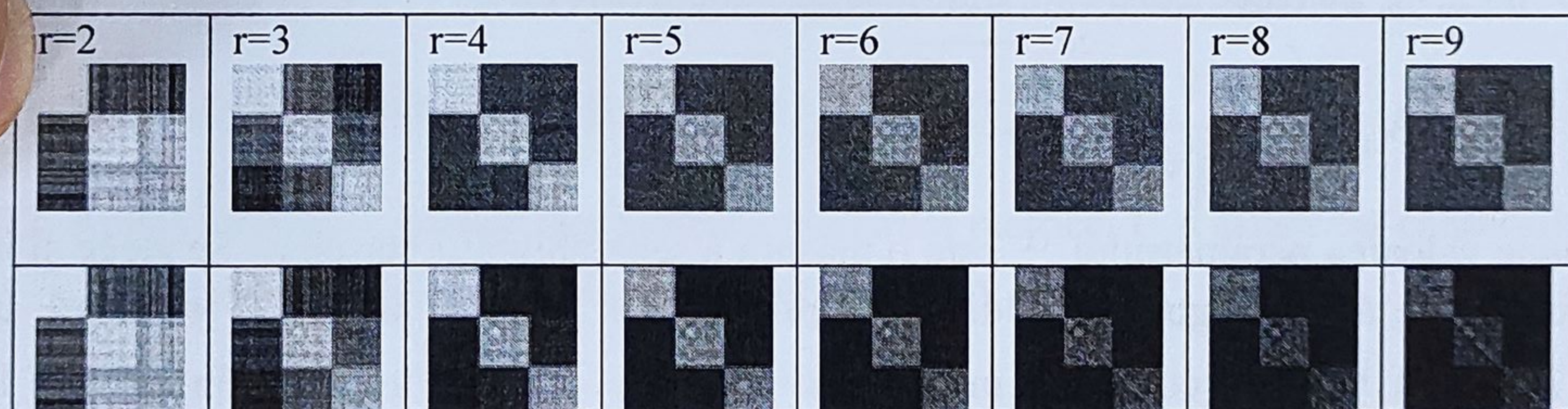


Fig. 9. Mișcări relaționate. Rândul de sus: matricele $A(r)$; Rândul al doilea: matricele H pentru $r=2,9$ (după Kim și Agapito²²).

Fig. 9 ilustrează matricele $A(r)$ și H în situația în care obiectele au în comun unele din mișcările de rotație. În timp ce blocurile ne-diagonale ale matricelor $A(r)$ sunt ne-nule cele ale matricelor H își păstrează proprietățile din mișcarea independentă, constituind un instrument mai potrivit pentru segmentarea mișcării.

O problemă este alegerea dimensiunii maxime D din relațiile (74), (75). În algoritmul de segmentare propus de Kim și Agapito este pus în evidență faptul că D este calculat astfel încât maximile din histogramele matricelor H să fie distanțate la

²⁰ Idem nota 14.

²¹ Jae-Hak Kim, Lourdes Agapito, „Motion Segmentation using the Hadamard Product and Spectral Clustering” în *Proceedings of the 2009 international conference on Motion and video computing* (WMVC 2009), December 2009, Snowbird, Utah, USA, pp. 126-133.

²² Idem nota 16.

cel puțin o valoare t_1 stabilită anterior iar diferența dintre elementele din jurul maximelor să fie mai mare decât o valoare de prag t_2 .

Algoritmul de segmentare Kim-Agapito este:

1. Determină matricea V din descompunerea SVD a matricei măsurătorilor W ;
2. $D=2$;
3. Pentru $r=2, N$:
 - 3.1. Calculează $Q(r)=V(r)V(r)t$; $A(r) = e^{Q(r)}$; $H_{ij} = H_{ij}A_{ij}/|A|$, cu $|A|$ norma matricei A ;
 - 3.2. Determină centriile c_1 și c_2 din histograma lui H ;
 - 3.3. Calculează $h_1=\text{histc}(A, c_1)$, $h_2=\text{histc}(A, c_2)$ numărul elementelor din A din vecinătatea fiecărui centroid c_i ;
 - 3.4. $D=D+1$;
 - 3.4. Continuă bucla atâta vreme cât: $|c_1 - c_2| < t_1$ sau $|h_1 - h_2| < t_2$.
3. Segmentează mișcarea utilizând matricea finală H : $H_{ij} = \prod_{r=2}^D A_{ij}(r)$.

Lauer și Schnorr au propus²³ ca alegerea dimensiunii D să se facă în felul următor. Se pleacă de la ideea că valoarea optimă pentru clusterizarea în n sub-spății d -dimensionale este $D=nd+1$. În cazul segmentării mișcării d variază între $d_{\min}=1$ și $d_{\max}=4$. Se descompune matricea măsurătorilor: $W = U \Sigma V^t$ și se inițializează $D=nd_{\min}+1$. Se calculează apoi matricea de similaritate A :

$$A_{ij} = \left(\frac{\mathbf{v}_i^t \mathbf{v}_j}{\|\mathbf{v}_i\| \|\mathbf{v}_j\|} \right)^{2\alpha} \quad (76)$$

unde valoarea parametrului α este o măsură a separabilității claselor. Se alege și în funcție de nivelul zgomotului din imagini. Lauer și Schnorr lucrează²⁴ cu $\alpha=4$. În (76) \mathbf{v}_i și \mathbf{v}_j sunt coloanele matricei $X=[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_D]$ formată din primele D coloane ale matricei V din descompunerea SVD a matricei măsurătorilor. Se calculează apoi matricea L :

$$L = D^{-1/2} A D^{-1/2} \quad (77)$$

unde D este o matrice diagonală construită după regula: $D_{ii} = \sum_{j=1}^N A_{ij}$.

Se evaluează apoi o măsură a separabilității clusterilor:

$$r_D = \frac{\lambda_n - \lambda_{n+1}}{\lambda_{n-1} - \lambda_n} \quad (78)$$

unde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{n+1}$ sunt cele mai mari $n+1$ valori proprii ale matricei L .

²³ Fabien Lauer, Christoph Schnorr, „Spectral clustering of linear subspaces for motion segmentation” în *The 12th IEEE International Conference on Computer Vision (ICCV'09)*, 2009, pp. 678-675.

²⁴ *Idem* nota 18.

Se incrementează valoarea curentă a lui D , $D=D+1$, și se reia procedeul de mai sus până când $D=nd_{\max}+1$.

Se alege apoi drept valoare optimă pentru $D=\max r_k$ și se aplică pentru această valoare algoritmul de clusterizare propus de Ng, Jordan și Weiss²⁵.

În concluzie, algoritmul Lauer-Schnorr este:

1. Determină matricea V din descompunerea SVD a matricei măsurătorilor W .

2. $D=nd_{\min}+1$.

3. Calculează matricea A conform (76).

4. Determină matricea L din (77) și primele ei $n+1$ valori proprii

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{n+1}.$$

5. $r_D = \frac{\lambda_n - \lambda_{n+1}}{\lambda_{n-1} - \lambda_n}.$

6. Dacă $D=nd_{\max}+1$ continuă cu pasul 7, dacă nu reia procesul de la pasul 3.

7. $D=\max r_k$; Construiește matricea $X=[v_1, v_2, \dots, v_D]$.

8. Construiește matricea A astfel: $A_{ij}=\exp(-\|v_i-v_j\|^2/2\sigma^2)$ dacă $i \neq j$, $A_{ii}=0$, cu $i,j=1,\dots,D$.

9. Construiește matricea diagonală D : $D_{ii} = \sum_{j=1}^N A_{ij}$ și matricea $L = D^{-1/2}AD^{-1/2}$.

10. Determină primii D vectori proprii ai matricei L și formează matricea U .

11. Normalizează rândurile lui U : $U_{ij} = U_{ij} / \sqrt{\sum_{j=1}^D U_{ij}^2}.$

12. Aplică metoda de clusterizare k-means rândurilor matricei U și asignează cluster-ului corespunzător fiecare coloană a matricei $X=[v_1, v_2, \dots, v_D]$.

2.2. Reconstrucție proiectivă

Vom discuta mai întâi un algoritm de reconstrucție proiectivă și apoi vom analiza un algoritm care tratează global, la nivelul întregului set de corespondențe, atât reconstrucția modelelor camerelor și a punctelor din spațiul proiectiv cât și calculul parametrilor $\lambda_{i,j}$.

Plecând de la ecuația exactă a proiecției x a unui punct X din spațiul proiectiv în planul imaginii camerei P :

$$\lambda x = PX \tag{79}$$

vom nota cu:

²⁵ Andrew Y. Ng, Michael I. Jordan, Yair Weiss, „On spectral clustering: Analysis and an algorithm” în *Advances in Neural Information Processing Systems* (NIPS'02), MIT Press, 2002, vol. 14, pp. 849–856.

$$\mathbf{x}_i = \begin{bmatrix} x_{i,1} & x_{i,2} & \cdot & \cdot & \cdot & x_{i,n} \\ y_{i,1} & y_{i,2} & \cdot & \cdot & \cdot & y_{i,n} \\ 1 & 1 & \cdot & \cdot & \cdot & 1 \end{bmatrix} \quad i = 1, \dots, m \quad (80)$$

matricea celor n pixeli corespondenți din imaginea i .

De asemenea, vom nota cu \mathbf{X} matricea celor n puncte din spațiul proiectiv vizualizate în cele m camere:

$$\mathbf{X} = \begin{bmatrix} X_1 & X_2 & X_3 & \cdot & \cdot & X_n \\ Y_1 & Y_2 & Y_3 & \cdot & \cdot & Y_n \\ Z_1 & Z_2 & Z_3 & \cdot & \cdot & Z_n \\ 1 & 1 & 1 & \cdot & \cdot & 1 \end{bmatrix} \quad (81)$$

și cu $\Lambda_i = \text{diag}(\lambda_{i,1}, \lambda_{i,2}, \dots, \lambda_{i,n})$ distanțele $\lambda_{i,j}$ corespunzătoare pixelului i în camera j .

Folosind notațiile de mai sus ecuațiile (79) pentru cele m camere sunt:

$$\mathbf{x}_i \Lambda_i = \mathbf{P}_i \mathbf{X} \quad (82)$$

Sistemul de ecuații (82) este așadar:

$$\begin{bmatrix} \mathbf{x}_1 \Lambda_1 \\ \mathbf{x}_2 \Lambda_2 \\ \mathbf{x}_3 \Lambda_3 \\ \cdot \\ \cdot \\ \mathbf{x}_m \Lambda_m \end{bmatrix} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \mathbf{P}_3 \\ \cdot \\ \cdot \\ \mathbf{P}_m \end{bmatrix} \mathbf{X} = \mathbf{P} \mathbf{X} \quad (83)$$

Notând subspațiul generat de rândurile matricelor \mathbf{x}_i cu D_i și cu D spațiul generat de rândurile matricei \mathbf{X} relațiile (82), (83) implică $\Lambda_i D_i \subset D$, respectiv $\Lambda_1 D_1 + \Lambda_2 D_2 + \dots + \Lambda_m D_m \subseteq D$, unde $S_1 + S_2 = \{x+y; x \in S_1, y \in S_2\}$.

În continuare, notând $\mathbf{x}_i \Lambda_i = \hat{\mathbf{x}}_i$, multiplicând fiecare membru al ecuației (83) cu transpusul său și făcând media pentru totalul imaginilor obținem:

$$\frac{1}{m} \sum_{i=1}^m \hat{\mathbf{x}}_i^t \hat{\mathbf{x}}_i = \mathbf{X}^t \left(\frac{1}{m} \sum_{i=1}^m \mathbf{P}_i^t \mathbf{P}_i \right) \mathbf{X} = \mathbf{X}^t \frac{1}{m} \mathbf{P}^t \mathbf{P} \mathbf{X} \quad (84)$$

Fie matricele $\mathbf{T}_i = \hat{\mathbf{x}}_i^t (\hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t)^{-1} \hat{\mathbf{x}}_i$ și $\mathbf{T} = \frac{1}{m} \sum_{i=1}^m \mathbf{T}_i$. Matricea \mathbf{T} este independentă de baza aleasă pentru reprezentarea vectorilor \mathbf{x}_i :

$$(\mathbf{A} \hat{\mathbf{x}}_i)^t (\mathbf{A} \hat{\mathbf{x}}_i (\mathbf{A} \hat{\mathbf{x}}_i)^t)^{-1} \mathbf{A} \hat{\mathbf{x}}_i = \hat{\mathbf{x}}_i^t \mathbf{A}^t (\mathbf{A}^t)^{-1} (\hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t)^{-1} \mathbf{A}^{-1} \mathbf{A} \hat{\mathbf{x}}_i = \hat{\mathbf{x}}_i^t (\hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t)^{-1} \hat{\mathbf{x}}_i = \mathbf{T}_i \quad (85)$$

Relația $\Lambda_1 D_1 + \Lambda_2 D_2 + \dots + \Lambda_m D_m \subseteq D$ implică de fapt că $\text{rang}(\mathbf{T}) \leq 4$, întrucât $\text{rang}(D) = 4$. Rezultă că, independent de sistemul de coordonate ales pentru fiecare cameră, a cincea valoare singulară din descompunerea SVD a matricei \mathbf{T} trebuie să fie nulă. Cum acest lucru, pentru un set de măsurători afectate de zgomot, nu se întâmplă Berthilsson, Heyden și Sparr²⁶ propun următoarea metodă de reconstrucție proiectivă, pornind de la setul de n corespondențe din primele două imagini din setul de m imagini:

1. Fie variabilele $q_{i,j} = \frac{\lambda_{i,j}}{\lambda_{1,j}}$, inițializate pentru început cu 1.
2. Se determină $\mathbf{T} = \frac{1}{m} \sum_{i=1}^m \hat{\mathbf{x}}_i^t (\hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^t)^{-1} \hat{\mathbf{x}}_i$.
3. Din descompunerea SVD $\mathbf{T} = \mathbf{U}^t \mathbf{S} \mathbf{U}$ se rețin matricea \mathbf{U} și a cincea valoare singulară, σ_5 . Dacă σ_5 este mai mică decât un anumit prag prestabilit și tot setul de imagini a fost procesat calculul coeficienților $\lambda_{i,j}$, matricelor \mathbf{X} și \mathbf{P}_i se încheie.
4. Fie \mathbf{X} matricea obținută din \mathbf{U} prin reținerea primelor patru rânduri.
5. Adaugă una sau mai multe imagini și completează șirul de corespondențe.
6. Pentru noul șir de corespondențe calculează Λ_i din $\mathbf{x}_i \Lambda_i = \mathbf{P}_i \mathbf{X}$ în felul următor: fie $\mathbf{T}_D = \mathbf{X}^t (\mathbf{X} \mathbf{X}^t)^{-1} \mathbf{X}$ și $\mathbf{Q}_D = \mathbf{I} - \mathbf{T}_D$. Relația $\mathbf{x}_i \Lambda_i = \mathbf{P}_i \mathbf{X}$ este echivalentă cu $\mathbf{Q}_D (\mathbf{x}_i \Lambda_i)^t = \mathbf{0}$, ceea ce înseamnă $3m$ ecuații lineare având m necunoscute (Λ_i). Soluția rezultă din optimizarea: $\min_{\|\Lambda_i\|=1} \sum_{j=1}^3 \|\mathbf{Q}_D \Lambda_i \mathbf{x}_i^j\|^2$ unde \mathbf{x}_i^j , $j=1,2,3$, sunt trei coloane ale matricei (80) considerate în această relație drept rânduri.
7. Determină $q_{i,j} = \frac{\lambda_{i,j}}{\lambda_{1,j}}$, calculează $\hat{\mathbf{x}}_i$ și continuă cu pasul 2.

Avantajul algoritmului lui Berthilsson, Heyden și Sparr²⁷ în raport cu metoda lui Heyden²⁸ constă în faptul că atunci când este nevoie doar de reconstrucția \mathbf{X} convergența este mai rapidă. În acest caz pasul 3 se modifică în sensul că procesarea se încheie atunci când valoarea singulară σ_5 este sub un prag prestabilit.

În continuare vom analiza algoritmul propus de Mahamud și Herbert²⁹ care spre deosebire de metoda lui Sturm și Triggs³⁰ tratează global, la nivelul întregului

²⁶ Richard Berthilsson, Anders Heyden, Gunnar Sparr, „Recursive Structure and Motion from Image Sequences using Shape and Depth Spaces” în *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 1997, pp. 444-449.

²⁷ Idem nota 25.

²⁸ Anders Heyden, „Reconstruction from multiple images by means of using relative depths” în *International Journal of Computer Vision*, 24(2), 1997, pp. 155-161.

²⁹ Shzjan Mahamud, Martial Herbert, „Iterative projective reconstruction from multiple views”, în *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00)*, June, 2000, pp. 430-437.

set de corespondențe, atât reconstrucția modelelor camerelor și a punctelor din spațiul proiectiv cât și calculul parametrilor $\lambda_{i,j}$. Sturm și Triggs propun³¹ determinarea acestora din perechi de imagini, ceea ce ar conduce la acumularea erorilor de calcul, în timp ce în primul caz³² distanțele $\lambda_{i,j}$ sunt calculate folosind întreg setul de corespondențe odată cu calculul matricelor camerelor și reconstrucției proiective. În plus algoritmul lui Sturm și Triggs este mai laborios întrucât pentru fiecare pereche de imagini trebuie determinate matricea fundamentală și epipolii. Metoda lui Mahamud și Herbert³³ calculează distanțele $\lambda_{i,j}$ simultan cu reconstrucția proiectivă (\mathbf{P} și \mathbf{X} din relațiile de mai sus) optimizând un criteriu derivat dintr-o condiție îndeplinită de toate cele m imagini ale aceluiași punct din spațiul proiectiv. Metoda este asemănătoare celei propuse de Berthilsson, Heyden și Sparr³⁴, cazul în care se determină reconstrucția proiectivă în totalitate, însă normalizarea datelor este diferită, formulată mai natural, ceea ce conduce la o altă problemă de optimizare.

În cele ce urmează vom arăta că dacă matricea \mathbf{P} din relația (83) este cunoscută parametrii $\lambda_{i,j}$ corespunzători imaginilor \mathbf{x}_{ij} ale un punct \mathbf{X}_j din spațiul proiectiv în toate cele $i=1,m$ imagini pot fi deduși rezolvând o problemă de vector propriu derivată dintr-o condiție de subspațiu.

Fie \mathbf{s}_j vectorul coloană rezultat din proiecția punctului \mathbf{X}_j în cele m imagini:

$$\mathbf{s}_j = [\lambda_{ij} \mathbf{x}_{ij}] = \mathbf{P} \mathbf{X}_j \quad (86)$$

Din relația de mai sus rezultă că dacă matricea \mathbf{P} este cunoscută se poate verifica dacă pixelii \mathbf{x}_{ij} rezultați din măsurătorile în cele m imagini sunt într-adevăr imaginile aceluiași punct \mathbf{X}_j atunci când putem calcula parametrii $\lambda_{i,j}$ și punctul \mathbf{X}_j .

Din expresia (86) se vede că vectorii \mathbf{s}_j aparțin unui subspațiu generat de combinații ale coloanelor matricei \mathbf{P} având coeficienții dați de vectorul \mathbf{X}_j . Condiția de subspațiu din care va fi derivată relația de calcul pentru $\lambda_{i,j}$ conduce la faptul că reziduul proiecției vectorului \mathbf{s}_j pe spațiul coloanelor lui \mathbf{P} este nul:

$$R_j(\lambda_j) = \frac{|(\mathbf{P}\mathbf{P}^+ - \mathbf{I})\mathbf{s}_j|^2}{|\mathbf{s}_j|^2} = 0 \quad (87)$$

unde $\lambda_j = (\lambda_{1,j}, \lambda_{2,j}, \dots, \lambda_{m,j})^t$ iar \mathbf{P}^+ este pseudo-inversa matricei \mathbf{P} .

De notat că reziduul R_j este normalizat pentru a se evita soluția trivială $\mathbf{s}_j = \mathbf{0}$. Expresia de mai sus se simplifică dacă se alege o bază ortonormală pentru spațiul coloa-

³⁰ Peter Sturm, Bill Triggs, „A Factorization Based Algorithm for Multi-Image Projective Structure and Motion” în *Proceedings of 4th European Conference on Computer Vision*, Cambridge, England, April 1996, pp. 709-720

³¹ *Idem* nota 25.

³² *Idem* nota 24.

³³ *Idem* nota 24.

³⁴ *Idem* nota 22.

nelor matricei \mathbf{P} . Fie \mathbf{U} matricea 3×4 ale cărei patru coloane constituie o bază ortonormală a spațiului coloanelor lui \mathbf{P} . Din cauză că $\mathbf{U}^+ = \mathbf{U}^t$ expresia reziduului se transformă în:

$$G_j(\lambda_j) = 1 - R_j(\lambda_j) = \frac{\mathbf{s}_j^t \mathbf{U} \mathbf{U}^t \mathbf{s}_j}{\mathbf{s}_j^t \mathbf{s}_j} = 1 \quad (88)$$

Punând în evidență necunoscutele $\lambda_{i,j}$ și corespondențele \mathbf{x}_{ij} avem $\mathbf{s}_j^t \mathbf{U} = \lambda_j^t \mathbf{A}_j$, unde rândul i al matricei \mathbf{A}_j este dat de $\mathbf{q}_{ij}^t \mathbf{U}_i$, cu \mathbf{U}_i matricea 3×4 formată din rândurile $3i, 3i+1, 3i+2$ ale matricei \mathbf{U} . Prin urmare:

$$G_j(\lambda_j) = \frac{\lambda_j^t \mathbf{A}_j \mathbf{A}_j^t \lambda_j}{\lambda_j^t \mathbf{B}_j \lambda_j} = 1 \quad (89)$$

unde matricea diagonală \mathbf{B}_j are elementul (i,i) de forma $\mathbf{x}_{ij}^t \mathbf{x}_{ij}$.

Câteva proprietăți ale funcției $G_j(\lambda_j)$:

1. Deoarece reziduul $R_j(\lambda_j) = 1 - G_j(\lambda_j) \geq 0$ rezultă că $G_j(\lambda_j)$ are valoarea maximă 1.
2. $G_j(\lambda_j)$ atinge valoarea maximă 1 pentru o anumită valoare a vectorului λ_j dacă într-adevăr pixelii \mathbf{x}_{ij} sunt imaginile aceluiasi punct \mathbf{X}_j .
3. Maximul lui $G_j(\lambda_j)$ este dat de cea mai mare valoare proprie μ din ecuația:

$$\mathbf{A}_j \mathbf{A}_j^t \lambda = \mu \mathbf{B}_j \lambda.$$

Din cele de mai sus rezultă că pentru a verifica faptul că pixelii \mathbf{x}_{ij} sunt imaginile punctului \mathbf{X}_j în cele m imagini trebuie ca valoarea proprie maximă dată de relația de mai sus să aibă valoarea 1. Vectorul λ_j căutat rezultă din vectorul propriu corespunzător valorii proprii maxime.

Din cele de mai sus rezultă că pornind de la corespondențele cunoscute \mathbf{x}_{ij} dacă am ști parametrii $\lambda_{i,j}$, matricele \mathbf{P}_i și reconstrucția \mathbf{X}_j se pot obține prin metodele de factorizare deja discutate. Reciproc, dacă am cunoaște matricea \mathbf{P} , folosind condiția de subspațiu discutată mai sus am putea determina $\lambda_{i,j}$. Folosind dependența circulară dintre necunoscutele \mathbf{P}_i și $\lambda_{i,j}$ se conturează următoarea metodă de reconstrucție proiectivă: pornind de la valorile inițiale $\lambda_{i,j} = 1$ se determină matricea \mathbf{P} , printr-una din metodele de factorizare folosite în cazul reconstrucției afine. Apoi cu noua matrice \mathbf{P} aplicând condiția de subspațiu se determină noile valori $\lambda_{i,j}$. Teoretic în pasul k al procesului iterativ acestea rezultă din minimizarea relației:

$$R_j(\lambda_j, \mathbf{X}_j) = \frac{|\mathbf{U}^k \mathbf{X}_j - \mathbf{s}_j|^2}{|\mathbf{s}_j|^2} \quad (90)$$

unde U^k este baza ortonormată a spațiului coloanelor lui P^k , matricea P a matricelor camerelor calculată în pasul k . Valorile optime pentru X_j în expresia de mai sus au forma $X_j = (U^k)^t s_j^*$, unde s_j^* este valoarea optimă care minimizează $R_j(\lambda_j)$, ceea ce rezultă din:

$$\frac{\partial R_j(\lambda_j, X_j)}{\partial X_j} = 0 \quad (91)$$

Ultima observație face ca optimizarea reziduului $R_j(\lambda_j, X_j)$ să fie echivalentă cu maximizarea criteriului:

$$G_j^k(\lambda_j) = \frac{\lambda_j^t A_j^k A_j^{kt} \lambda_j}{\lambda_j^t B_j^k \lambda_j} = 1 \quad (92)$$

unde A_j^k și B_j^k au aceeași semnificație ca cele din expresia (89) doar că sunt calculate în pasul k al procesului iterativ, pornind de la baza ortonormată U^k .

Valorile $\lambda_{i,j}$ se determină din ecuația $A_j A_j^t \lambda = \mu B_j \lambda$ aflând vectorul propriu corespunzător valorii proprii maxime $\mu = 1$.

Cu noile valori $\lambda_{i,j}$ introduse în matricea măsurătorilor se determină prin factorizare următoarea bază ortonormată U^{k+1} și procesul continuă până când valorile $\lambda_{i,j}$ se stabilizează. Desigur, în fiecare pas calculul vectorului λ_j , aferent corespundențelor $x_{1j} \leftrightarrow x_{2j} \dots \leftrightarrow x_{nj}$, imaginile punctului X_j , se repetă pentru toate punctele din spațiul proiectiv $X_j, j=1, m$.

În continuare vom discuta convergența șirurilor $\lambda_j, j=1, m$. Vom nota cu λ matricea $[\lambda_{i,j}]$. Fie funcția obiectiv: $G^k(\lambda) = \frac{1}{n} \sum_{j=1}^n G_j^k(\lambda_j)$. Matricele A_j^k și B_j^k din expresia $G_j^k(\lambda_j)$ sunt calculate în pasul k al procesului iterativ, pornind de la baza ortonormată U^k . Întru-cât $G_j^k(\lambda_j) \leq 1$ funcția obiectiv va fi de asemenea subunitară: $G^k(\lambda) \leq 1$. Mahamud și Herbert au demonstrat³⁵ următoarea teoremă:

Teoremă Fie λ^k matricea λ la iterația k . Funcția obiectiv de mai sus converge monoton către un maxim:

$$G^{k+1}(\lambda^{k+1}) \geq G^k(\lambda^k) \quad (93)$$

Demonstrație La începutul iterației k dispunem de U^{k-1} și λ^{k-1} din iterația anterioară. Mai întâi este menținută constantă matricea U^{k-1} și este maximizată funcția obiectiv

³⁵ Shzjan Mahamud, Martial Herbert, „Iterative projective reconstruction from multiple views” în *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00)*, June, 2000, pp. 430–437.

$G^{k-1}(\lambda)$ pentru obținerea matricei λ^k . Prin urmare vom avea $G^{k-1}(\lambda^k) \geq G^{k-1}(\lambda^{k-1})$. În caz de egalitate procesul iterativ se încheie. Altfel, se determină U^k din factorizarea noii matrice a măsurătorilor construită din setul inițial de corespondențe și noii factori λ^k . Notăm noua matrice a măsurătorilor cu W^k . Coloanele j ale acesteia sunt normalizate de $((\lambda_j^k)^t B_j \lambda_j^k)$:

$$W_{ij}^k = \frac{\lambda_{ij}^k q_{ij}}{(\lambda_j^k)^t B_j \lambda_j^k} \quad (94)$$

Din descompunerea SVD a matricei W^k construim matricea U^k formată din cele patru coloane ale primei matrice a descompunerii corespunzătoare celor mai mari patru valori singulare. Rămâne să arătăm că $G^k(\lambda^k) \geq G^{k-1}(\lambda^k)$, prin tranzitivitate rezultând apoi ceea ce trebuia demonstrat: $G^k(\lambda^k) \geq G^{k-1}(\lambda^{k-1})$.

Proiecția unei matrice W pe un subspațiu U este dată de $\text{tr}(U^t W W^t U)$. Lema extensiei a lui Poincare arată că proiecția unei matrice W pe un subspațiu ortogonal U , r dimensional, este maximă atunci când subspațiul este generat de primii r vectori singulari U^* ai matricei W . Adică:

$$\max \text{tr}(U^t W W^t U) = \text{tr}(U^{*t} W W^t U^*) \quad (95)$$

Pentru $r=4$ se poate vedea că proiecția lui W^k pe $U=U^{k-1}$ nu este altceva decât $G^{k-1}(\lambda^k)$. Notând $U^* = U^k$ din descompunerea SVD a lui W^k rezultă $G^k(\lambda^k) \geq G^{k-1}(\lambda^k)$. În final avem: $G^k(\lambda^k) \geq G^{k-1}(\lambda^{k-1})$, ceea ce trebuia demonstrat.

În continuare vom analiza din alt punct de vedere problema rezolvării sistemului (83) pornind de la matricea celor n corespondențe din m imagini. Rescriem (83) în forma:

$$\begin{bmatrix} \lambda_{11} \mathbf{x}_{11} & \lambda_{12} \mathbf{x}_{12} & \cdot & \cdot & \lambda_{1n} \mathbf{x}_{1n} \\ \lambda_{21} \mathbf{x}_{21} & \lambda_{22} \mathbf{x}_{22} & \cdot & \cdot & \lambda_{2n} \mathbf{x}_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda_{m1} \mathbf{x}_{m1} & \lambda_{m2} \mathbf{x}_{m2} & \cdot & \cdot & \lambda_{mn} \mathbf{x}_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \cdot \\ \cdot \\ \mathbf{P}_m \end{bmatrix} [\mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdot \quad \cdot \quad \mathbf{X}_n] \quad (96)$$

Ideal ar fi dacă s-ar putea folosi ca în cazul afin optimizarea care minimizează distanța dintre pixelii corespondenți detectați în urma procesării celor m imagini și pixelii estimați pornind de la parametrii λ_{ij} , \mathbf{X}_j și \mathbf{P}_i :

$$E = \frac{1}{mn} \sum_{i,j} \left| \mathbf{x}_{ij} - \frac{1}{\lambda_{ij}} \mathbf{P}_i \mathbf{X}_j \right|^2 \quad (97)$$

Din păcate minimizarea lui E în raport cu λ_{ij} , \mathbf{P}_i și \mathbf{X}_j conduce la sisteme neliniare greu de rezolvat. Folosind însă notațiile de mai sus

$\mathbf{s}_j = (\lambda_{1j}\mathbf{x}_{1j}, \lambda_{2j}\mathbf{x}_{2j}, \dots, \lambda_{mj}\mathbf{x}_{mj})^t$ și $\boldsymbol{\lambda}_j = (\lambda_{1j}, \lambda_{2j}, \dots, \lambda_{mj})^t$ vom minimiza $E = \frac{1}{mn} \sum_j E_j$ cu $E_j = \frac{1}{|\mathbf{s}_j|^2} \|\mathbf{s}_j - \mathbf{P}\mathbf{X}_j\|^2$, ceea ce înseamnă minimizarea fiecărui E_j

în raport cu $\boldsymbol{\lambda}_j$, \mathbf{P} și \mathbf{X}_j . Adică minimizarea distanțelor dintre proiecțiile punctului \mathbf{X}_j în toate cele m camere și imaginile sale \mathbf{x}_{ij} scalate de parametrii λ_{ij} . Procedăm ca mai sus la minimizări succesive, mai întâi considerând \mathbf{P} fix și optimizând E_j în raport cu $\boldsymbol{\lambda}_j$ și \mathbf{X}_j iar apoi cu noile valori $\boldsymbol{\lambda}_j$ minimizăm E în raport cu \mathbf{P} și \mathbf{X} prin factorizare.

Ca și în cazul afin minimizarea lui E_j în raport cu \mathbf{X}_j conduce la faptul că optimul lui E_j se obține pentru $\mathbf{X}_j = \mathbf{P}^+ \mathbf{s}_j$, rezultat din $\frac{\partial E_j}{\partial \mathbf{X}_j} = 0$, unde

$$\mathbf{P}^+ = (\mathbf{P}^t \mathbf{P})^{-1} \mathbf{P}^t:$$

$$E_j = \frac{1}{|\mathbf{s}_j|^2} \|(\mathbf{I} - \mathbf{P}\mathbf{P}^+) \mathbf{s}_j\|^2 \quad (98)$$

Cum \mathbf{P} este o matrice $3m \times 4$ descompunerea SVD va avea forma $\mathbf{P} = \mathbf{U}\mathbf{W}\mathbf{V}^t$, cu \mathbf{U} o matrice $3m \times 4$ având coloanele ortogonale, \mathbf{W} o matrice diagonală nesingulară 4×4 și \mathbf{V}^t o matrice ortogonală 4×4 . Pseudo-inversa matricei \mathbf{P} a matricelor camerelor este așadar $\mathbf{P}^+ = \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^t$, ceea ce transformă relația de mai sus în:

$$E_j = \frac{1}{|\mathbf{s}_j|^2} \|(\mathbf{I} - \mathbf{U}\mathbf{U}^t) \mathbf{s}_j\|^2 = \frac{1}{|\mathbf{s}_j|^2} \left[\|\mathbf{s}_j\|^2 - \mathbf{s}_j^t (\mathbf{U}\mathbf{U}^t) \mathbf{s}_j \right] = 1 - \frac{\mathbf{s}_j^t (\mathbf{U}\mathbf{U}^t) \mathbf{s}_j}{|\mathbf{s}_j|^2} \quad (99)$$

Ceea ce înseamnă că minimizarea criteriului E_j se transformă în maximizarea expresiei $\frac{\mathbf{s}_j^t (\mathbf{U}\mathbf{U}^t) \mathbf{s}_j}{|\mathbf{s}_j|^2}$ în raport cu $\boldsymbol{\lambda}_j$. Deoarece $\mathbf{s}_j^t = \boldsymbol{\lambda}_j^t \mathbf{Q}_j$, unde

$$\mathbf{Q}_j = \begin{bmatrix} \mathbf{x}_{1j}^t & 0 & \cdot & \cdot & 0 \\ 0 & \mathbf{x}_{2j}^t & & & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \mathbf{x}_{nj}^t \end{bmatrix}, \text{ în final trebuie optimizată expresia: } \frac{\boldsymbol{\lambda}_j^t \mathbf{A}_j \boldsymbol{\lambda}_j}{\boldsymbol{\lambda}_j^t \mathbf{B}_j \boldsymbol{\lambda}_j} \quad (100)$$

unde $\mathbf{A}_j = \mathbf{Q}_j \mathbf{U}\mathbf{U}^t \mathbf{Q}_j^t$ iar $\mathbf{B}_j = \mathbf{Q}_j \mathbf{Q}_j^t$. Prin urmare trebuie maximizat criteriul $G_j^k(\boldsymbol{\lambda}_j)$ de mai sus.

Așa cum s-a arătat, cu noile valori $\lambda_{i,j}$ introduse în matricea măsurătorilor se determină prin factorizare următoarea bază ortonormată U^{k+1} și procesul continuă până când valorile $\lambda_{i,j}$ se stabilizează.

Sintetizând, algoritmul de reconstrucție proiectivă discutat mai sus are următorii pași:

1. Pornind de la valorile inițiale $\lambda_{i,j}=1$ repetă până la convergență pașii:
2. Normalizează fiecare coloana a matricei

$$D = \begin{bmatrix} \lambda_{11} \mathbf{x}_{11} & \lambda_{12} \mathbf{x}_{12} & \cdot & \cdot & \lambda_{1n} \mathbf{x}_{1n} \\ \lambda_{21} \mathbf{x}_{21} & \lambda_{22} \mathbf{x}_{22} & \cdot & \cdot & \lambda_{2n} \mathbf{x}_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda_{m1} \mathbf{x}_{m1} & \lambda_{m2} \mathbf{x}_{m2} & \cdot & \cdot & \lambda_{mn} \mathbf{x}_{mn} \end{bmatrix};$$

3. Din descompunerea SVD a matricei $D=UWV^t$ formează matricea U_4 reținând primele 4 coloane ale matricei U .
4. Pentru toate corespondențele $j=1,n$ repetă pașii:
 - 4.1. calculează $A_j=Q_j U_4 U_4^t Q_j^t$ și $B_j=Q_j Q_j^t$;
 - 4.2. calculează vectorul λ_j ;
 - 4.3. actualizează coloana corespunzătoare a matricei D cu noile valori λ_j .

3. Tehnici de urmărire a unui corp rigid; Tehnici de calcul pentru estimarea 3D a mișcării

În cele ce urmează vom analiza un sistem de urmărire a unui robot mobil care se deplasează în mediu natural. Pe robotul mobil este montat un sistem trinocular de camere video care captează imagini ale mediului. În urma procesării secvenței de imagini achiziționate pe parcurs în final se determină traiectoria robotului într-un sistem de referință extern. Procesul de urmărire constă din pașii:

1. Extragerea din imaginea curentă a trăsăturilor care vor fi urmărite în cadrul următor.
2. Calculul coordonatelor 3D ale trăsăturilor extrase prin tehnici stereo.
3. Urmărirea trăsăturilor 3D în secvența de imagini.
4. Estimarea mișcării: Calculul poziției robotului într-un sistem de referință extern.
5. Corecția poziției robotului și a coordonatelor 3D ale caracteristicilor scenei.

În continuare vom discuta în detaliu pașii de mai sus.

3.1. Extragerea trăsăturilor

Într-o primă etapă, Saeedi, Lawrence și Lowe au aplicat³⁶ algoritmul de

³⁶ Parvaneh Saeedi, Peter D. Lawrence, David G. Lowe, „3D Motion Tracking of a Mobile Robot in a Natural Environment” în *Proceedings of the 2000 IEEE International Conference on Robotics and Automation*, ICRA 2000, pp. 1682-168.

extragere de colțuri propus de Harris și Stephens³⁷, potrivit căruia colț este declarat orice pixel pentru care o funcție de răspuns, descrisă mai jos, notată cu R , are o valoare pozitivă semnificativă. Calculul lui R presupune definirea mai întâi a unui criteriu $E(x,y)$ calculat pentru deplasări mici ale unei ferestre circulare de interes, notată cu $W_{u,v}$, în jurul pixelului curent (u,v) :

$$E(x,y) = W_{u,v} |I_{x+u,y+v} - I_{u,v}|^2 \quad (101)$$

În relația (101) $I_{u,v}$ este intensitatea pixelului (u,v) din imaginea curentă iar x, y reprezintă deplasarea ferestrei circulare $W_{u,v} = e^{-\frac{u^2+v^2}{\sigma^2}}$.

În ipoteza unor deplasări mici ale ferestrei criteriul $E(x,y)$ poate fi aproximat ca urmare a dezvoltării în serie Taylor a termenului $I_{x+u,y+v} - I_{u,v}$ din ecuația (101):

$$E(x,y) = W_{u,v} |I_{x+u,y+v} - I_{u,v}|^2 = W_{u,v} |xX + yY + O^2(x^2, y^2)|^2 \approx \approx W_{u,v} |xX + yY|^2 \quad (102)$$

unde

$$X = I \otimes (-1, 0, 1) \approx \frac{\partial I}{\partial x}, \quad Y = I \otimes (-1, 0, 1)^t \approx \frac{\partial I}{\partial y} \quad (103)$$

Așadar $E(x,y)$ poate fi scris ca:

$$E(x,y) = Ax^2 + 2Cxy + By^2 \quad (104)$$

unde

$$A = X^2 \otimes W, \quad B = Y^2 \otimes W, \quad C = XY \otimes W \quad (105)$$

$$\text{În final } E(x,y) = \begin{bmatrix} x & y \end{bmatrix} \mathbf{M} \begin{bmatrix} x \\ y \end{bmatrix} \text{ cu } \mathbf{M} = \begin{bmatrix} A & C \\ C & B \end{bmatrix}.$$

Colț este declarat orice pixel pentru care $R = \text{Det}(\mathbf{M}) - K * (\text{Tr}(\mathbf{M}))^2$ are o valoare pozitivă semnificativă. În expresia lui R coeficientul K reprezintă raportul maxim al valorilor proprii ale matricei \mathbf{M} pentru situația în care funcția de răspuns R are valori pozitive. Pentru regiuni uniforme R are valori în jurul lui 0 iar pentru muchii valori negative.

Metoda de mai sus este eficientă în selecția colțurilor. Totuși în aplicații de timp real cum este cea descrisă, când imaginile trebuie procesate de la cadru la cadru, devine prohibitivă fiind consumatoare de timp.

³⁷ Chris Harris, Mike Stephens, „A combined corner and edge detector” în *Proceedings of 4th Alvey Vision Conference*, 1988, pp. 147-151.

O metodă mai rapidă decât cea a lui Harris și Stephens³⁸ a fost propusă de Saeedi, Lawrence și Lowe³⁹ și este aplicată pentru determinarea traiectoriei robotului mobil^{40, 41}.

Într-un prim pas imaginea inițială este binarizată. Mai întâi se aplică un filtru Gaussian. Pentru $\sigma = 0.8$ filtrul poate fi aproximat prin convoluții cu masca $[0.25, 0.5, 1, 0.5, 0.25]$. În acest fel cele patru înmulțiri presupuse de convoluție sunt înlocuite de patru operații de shifting. Apoi în imaginea rezultată, în fiecare pixel $I(i, j)$, este calculat Laplacianul potrivit formulei:

$$\frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \approx (I_{i-1,j} + I_{i,j-1} + I_{i+1,j} + I_{i,j+1} - 4I_{i,j}) \quad (106)$$

Imaginea binară este generată în funcție de semnul Laplacianului (106) în fiecare punct:

$$L(i, j) = \begin{cases} 1 & \text{daca } \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} > 0 \\ 0 & \text{altfel} \end{cases}$$

Apoi în fiecare pixel al imaginii binare rezultate este plasată o masca circulară cu diametrul de 7 pixeli. Valoarea binară a fiecărui pixel p din interiorul măștii este comparată cu cea a pixelului central p_0 :

$$C(p_0, p) = \begin{cases} 1 & \text{daca } L(p) = L(p_0) \\ 0 & \text{daca } L(p) \neq L(p_0) \end{cases} \quad (107)$$

unde $L(p)$ este valoarea binară a pixelului $p(x, y)$ din interiorul măștii circulare iar $L(p_0)$ este valoarea pixelului central.

La nivelul întregii măști w se calculează:

$$n(p_0) = \sum_w C(p_0, p) \quad (108)$$

În (108) $n(p_0)$ este aria zonei măștii unde semnul Laplacianului este același cu cel al Laplacianului punctului central.

Pixelul central P_0 este luat în considerare ca posibil colț dacă $n(p_0)$ este mai mic decât jumătate din pixelii ferestrei w , valoare pe care o vom nota cu t . Prin urmare un pixel este candidat colț dacă $R(p_0)$, definit mai jos, este pozitiv.

³⁸ Chris Harris, Mike Stephens, „A combined corner and edge detector” în *Proceedings of 4th Alvey Vision Conference*, 1988, pp. 147-151.

³⁹ Parvaneh Saeedi, David G. Lowe, Peter D. Lawrence: „An efficient binary corner detector” în *Proceedings of Seventh International Conference on Control, Automation, Robotics and Vision*, ICARCV 2002, pp. 338-343.

⁴⁰ Parvaneh Saeedi, David G. Lowe, Peter D. Lawrence, „3D localization and tracking in unknown environments” în *Proceedings of the 2003 IEEE International Conference on Robotics and Automation*, ICRA 2003, pp. 1297-1303.

⁴¹ Parvaneh Saeedi, Peter D. Lawrence, David G. Lowe, „Vision-based 3-D trajectory tracking for unknown environments” în *IEEE Transactions on Robotics*, Feb. 2006, vol. 22, pp. 119-136.

$$R(p_0) = \begin{cases} n(p_0) & \text{daca } n(p_0) < t \\ 0 & \text{altfel} \end{cases} \quad (109)$$

Apoi pentru fiecare pixel cu $R(p_0) > 0$ se calculează criteriul $G(p_0)$:

$$G(p_0) = \sqrt{(g(x_0))^2 + (g(y_0))^2} \quad (110)$$

unde

$$g(x_0) = \frac{\sum (x_0 - x)}{n(p_0)}, \quad g(y_0) = \frac{\sum (y_0 - y)}{n(p_0)} \quad (111)$$

Centrul de greutate $G(p_0)$ furnizând direcția colțului este folosit drept criteriu la filtrarea candidaților. Se aplică condiția:

$$G(p_0) > |r_g| \quad (112)$$

Un ultim criteriu de selecție a colțurilor constă în aplicarea unui criteriu luând în considerare derivata direcțională a candidaților. Candidații cu derivate mici direcționale sunt eliminați. Ultima condiție de colț este:

$$|I(p_0) - I(p)| > I_t \quad (113)$$

unde I_t este pragul pentru variația intensității.

3.2. Calculul coordonatelor 3D ale trăsăturilor extrase prin tehnici stereo

Stabilirea perechilor de trăsături în cele două camere ale fiecărui sistem stereo se face aplicând mai întâi criteriul sumei normalizate a pătratelor diferențelor pentru stabilirea unui set inițial de perechi. După care perechile stabilite inițial sunt triate prin metoda minimizării disparităților⁴², discutată pe scurt la sfârșitul paragrafului, pentru cazul unidimensional.

Odată stabilite caracteristicile punctuale imagistice de urmărit se determină punctele din scena reală ale căror imagini aceste caracteristici sunt, folosind fie un sistem stereo fie un sistem trinocular, ilustrat de fig.10. În sistemul stereo cele două camere sunt aliniate orizontal astfel încât epipolara corespunzătoare oricărui pixel dintr-o imagine să fie paralelă cu axa Ox a sistemului celeilalte imagini⁴³. În sistemul trinocular perechii de camere aliniată orizontal i se mai adaugă o cameră aliniată vertical cu camera din dreapta a sistemului stereo inițial, care devine camera centrală a sistemului trinocular. Deși reconstrucția este posibilă din doar două imagini cea de

⁴² Masatoshi Okutomi, Takeo Kanade, „A multiple-baseline stereo” în *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (PAMI), vol. 15(4), 1993, pp. 353-363.

⁴³ Florin Rotaru, Dan Gălea, Ion Păvăloi, Silviu Bejinariu, Cristina Niță, Ramona Luca, „Reconstrucție 3D din două imagini” în *Raport de cercetare*, noiembrie 2007, IIT Iași.

a treia cameră este introdusă pentru a elimina colțurile false, cele ale căror imagini nu se regăsesc în toate cele trei camere.

În figura 11 este ilustrat sistemul stereo format din camera din stânga și camera centrală a sistemului trinocular din fig. 10. În continuare vom nota cu f distanța focală a camerelor și cu B_h distanța dintre centrele camerelor, presupuse cunoscute în urma calibrării sistemului stereo. Coordonatele 3D al punctului $P(X,Y,Z)$ din scena 3D în funcție de imaginile sale din camera din stânga, $P_l(x_l, y_l)$, respectiv camerei centrale $P_c(x_c, y_c)$, ținând seama că $y_l = y_c$, sunt:

$$Z = f \frac{B_h}{x_c - x_l}$$

$$X = Z \frac{x_l}{f}$$

$$Y = Z \frac{y_l}{f}$$

(114)

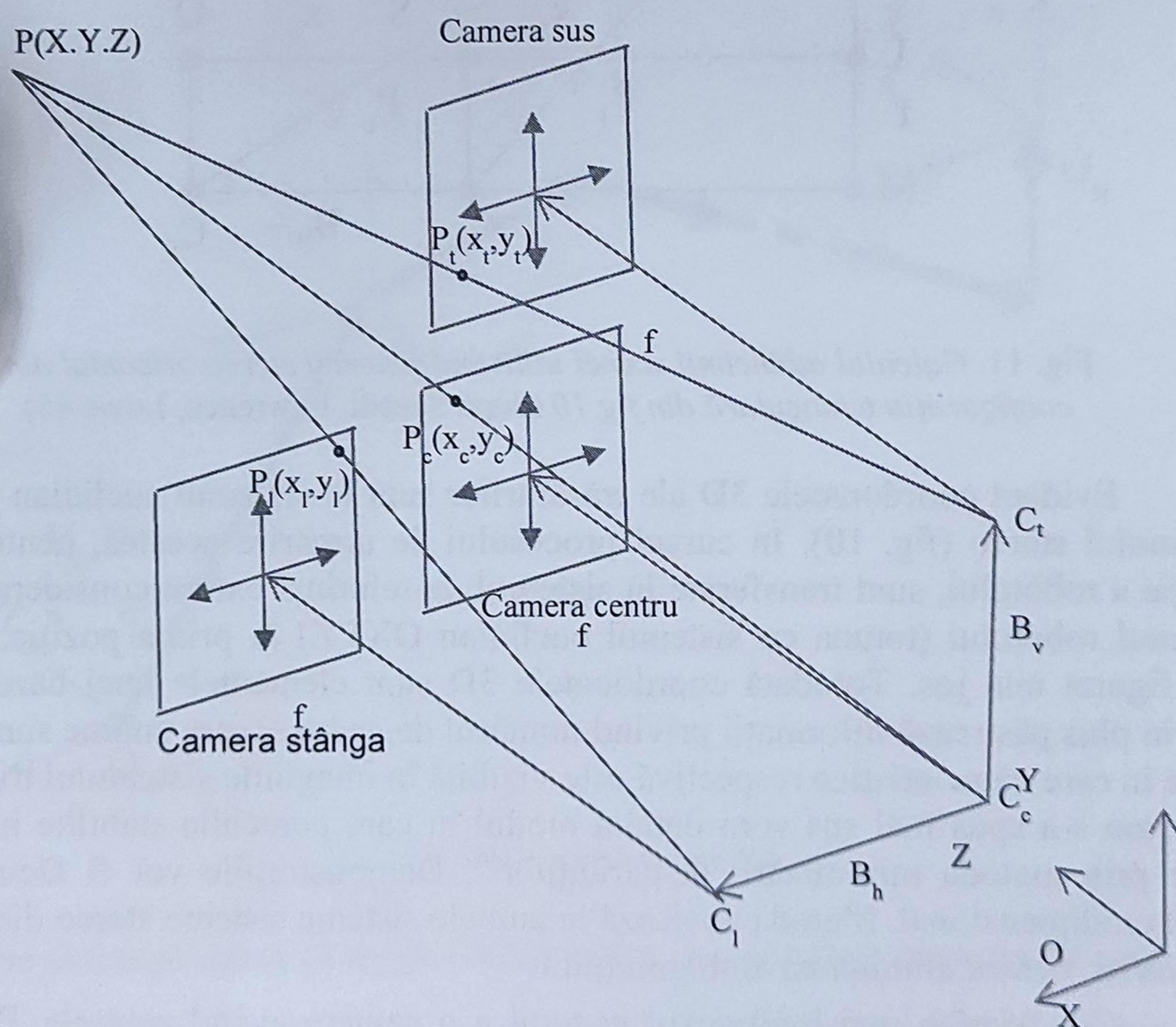


Fig. 10. Sistem trinocular de camere (după Takezawa, Herath și Dissanayake⁴⁴).

⁴⁴ Satoshi Takezawa, Mudiyanse Herath, Gamini Dissanayake, „SLAM in indoor environments with stereo vision” în *Proceedings of 2004 International Conference on Intelligent Robots and Systems*, (IROS 2004), vol. 2, pp. 1866-1871.

Prima din relațiile (114) rezultă din rapoartele generate de următoarele asemănări de triunghiuri din fig.11:

$$\Delta PLP_1 \sim \Delta PMC_1 \Rightarrow \frac{PL}{PM} = \frac{PP_1}{PC_1}, \frac{Z-f}{Z} = \frac{PP_1}{PC_1} \quad (115)$$

$$\Delta PP_1 P_c \sim \Delta PC_1 C_c \Rightarrow \frac{PP_1}{PC_1} = \frac{PP_1}{PC_1}, \frac{B_h - x_c + x_l}{B_h} = \frac{PP_1}{PC_1} \quad (116)$$

$$\frac{Z-f}{Z} = \frac{B_h - x_c + x_l}{B_h} \Rightarrow Z = f \frac{B_h}{x_c - x_l} \quad (117)$$

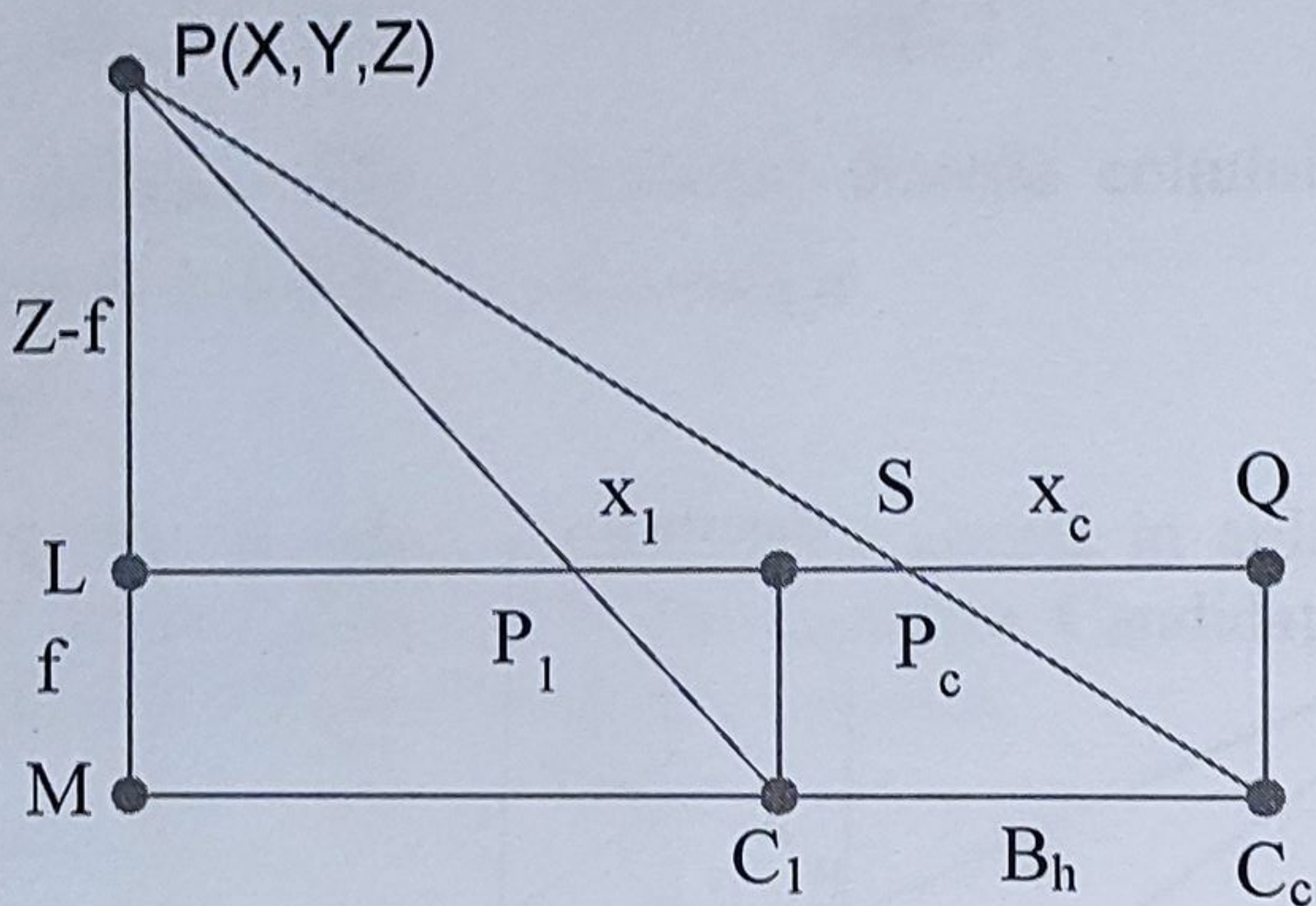


Fig. 11. *Calculul adâncimii scenei utilizând sistemul stereo orizontal din configurația trinoculară din fig.10 (după Saeedi, Lawrence, Lowe 45)*

Evident coordonatele 3D ale trăsăturilor sunt în sistemul euclidian OXYZ al sistemului stereo (fig. 10). În cursul procesului de urmărire acestea, pentru fiecare poziție a robotului, sunt transferate în sistemul de referință extern considerat ca fiind sistemul robotului (totuna cu sistemul euclidian OXYZ) în prima poziție, așa cum este figurat mai jos. Totodată coordonatele 3D sunt elementele unei baze de date, care în plus păstrează informații privind numărul de cadre și care anume sunt acestea, cadre în care caracteristica respectivă este vizibilă în imaginile sistemului trinocular. Așa cum s-a spus mai sus vom detalia modul în care perechile stabilite inițial sunt triate prin metoda minimizării disparităților⁴⁶. Demonstrațiile vor fi făcute pentru cazul unidimensional. Metoda apelează la ambele sisteme stereo din sistemul trinocular, pentru eliminarea ambiguităților.

Considerăm mai întâi cazul general a n camere având centrele P_0, \dots, P_n aliniate orizontal la distanțele B_1, \dots, B_n , în așa fel încât axele lor optice sunt

⁴⁵ Parvaneh Saeedi, Peter D. Lawrence, David G. Lowe, „3D Motion Tracking of a Mobile Robot in a Natural Environment” in *Proceedings of the 2000 IEEE International Conference on Robotics and Automation*, ICRA 2000, pp. 1682-168.

⁴⁶ Masatoshi Okutomi, Takeo Kanade, „A multiple-baseline stereo” in *IEEE Transactions on Pattern Analysis and Machine Intelligence* (PAMI), vol. 15(4), 1993, pp. 353-363.

perpendiculare pe dreapta centrelor camerelor. Notăm cu $f_0(x)$ și $f_i(x)$ imaginile în camerele P_0 și P_i ale aceluiași punct Z al scenei 3D, a cărei adâncime o vom nota cu z . Fie $d_{r(i)}$ diferența dintre coordonatele x ale imaginilor punctului Z . Așa cum s-a arătat mai sus relația dintre coordonata z a punctului Z și distanța B_i dintre centrele camerelor P_0 și P_i , considerând camerele calibrate este: $d_{r(i)} = f \frac{B_i}{z}$, unde f este distanța focală a primei camere. Vom modela nivelele de gri $f_0(x)$ și $f_i(x)$ ale imaginilor aceluiași punct din spațiu potrivit formulelor:

$$f_0(x) = f(x) + n_0(x)$$

$$f_i(x) = f(x - d_{r(i)}) + n_i(x)$$

unde $n_0(x)$, $n_i(x)$ reprezintă zgomot alb gaussian: $n_0(x), n_i(x) \sim N(0, \sigma_n^2)$.

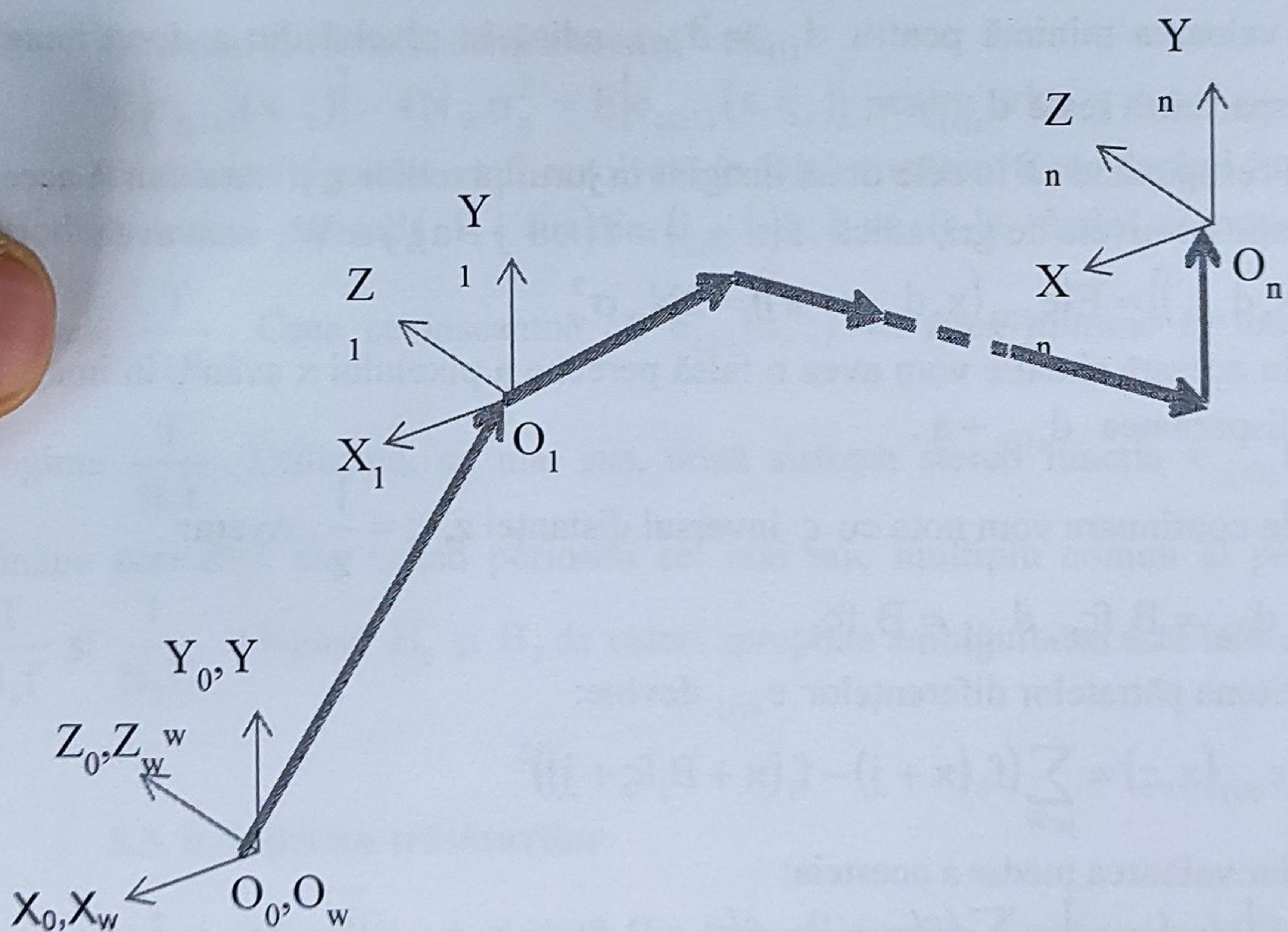


Fig. 12. Traectoria robotului.

Se definește suma pătratelor diferențelor $e_{d(i)}$ pentru o fereastră W centrată în poziția x a pixelului de intensitate $f_0(x)$ pentru posibila disparitate $d_{(i)}$ a pixelului pereche din a doua imagine ca fiind:

$$e_{d(i)}(x, d_{(i)}) = \sum_{j \in W} (f_0(x + j) - f_i(x + d_{(i)} + j))^2$$

Valoarea medie a lui $e_{d(i)}(x, d_{(i)})$ este:

$$\begin{aligned}
 E(e_{d(i)}(x, d_{(i)})) &= E \left[\sum_{j \in W} (f(x+j) - f(x+d_{(i)} - d_{r(i)} + j) + n_0(x+j) - n_1(x+d_{(i)} + j))^2 \right] = \\
 &= E \left[\sum_{j \in W} (f(x+j) - f(x+d_{(i)} - d_{r(i)} + j))^2 \right] + \\
 &E \left[\sum_{j \in W} (2(f(x+j) - f(x+d_{(i)} - d_{r(i)} + j))(n_0(x+j) - n_1(x+d_{(i)} + j))) \right] + \\
 &E \left[\sum_{j \in W} (n_0(x+j) - n_1(x+d_{(i)} + j))^2 \right] = \\
 &E \left[\sum_{j \in W} (f(x+j) - f(x+d_{(i)} - d_{r(i)} + j))^2 \right] + 2N_w \sigma_n^2
 \end{aligned}$$

unde N_w este numărul pixelilor din fereastra W . Este evident că funcția $e_{d(i)}(x, d_{(i)})$ va avea valoarea minimă pentru $d_{(i)} = d_{r(i)}$, adică în pixelul din a doua imagine având disparitatea reală $d_{r(i)}$.

Presupunând că în cele două imagini în jurul pixelilor x și $x+a$ există aceeași configurație de nivele de gri, adică $f(x+j) = f(x+j+a)$, $j \in W$, vom avea:

$$E(e_{d(i)}(x, d_{r(i)})) = E(e_{d(i)}(x, d_{r(i)} + a)) = 2N_w \sigma_n^2$$

În această situație vom avea o falsă pereche a pixelului x având, în imaginea a doua, disparitatea $d_{r(i)} + a$.

În continuare vom nota cu ς inversul distanței z , $\varsigma = \frac{1}{z}$. Avem:

$$d_{(i)} = B_i f \varsigma, \quad d_{r(i)} = B_i f \varsigma_r.$$

Suma pătratelor diferențelor $e_{d(i)}$ devine:

$$e_{\varsigma(i)}(x, \varsigma) = \sum_{j \in W} (f_0(x+j) - f_i(x + B_i f \varsigma + j))^2$$

Iar valoarea medie a acesteia:

$$E[e_{\varsigma(i)}(x, \varsigma)] = \sum_{j \in W} (f(x+j) - f(x + B_i f(\varsigma - \varsigma_r) + j))^2 + 2N_w \sigma_n^2$$

Se definește o nouă funcție de evaluare ca fiind suma funcțiilor $e_{\varsigma(i)}(x, \varsigma)$ pentru n sisteme stereo:

$$e_{\varsigma(12...n)}(x, \varsigma) = \sum_{i=1}^n e_{\varsigma(i)}(x, \varsigma)$$

Valoarea medie a acesteia este:

$$E[e_{\zeta(12...n)}(x, \zeta)] = \sum_{i=1}^n E[e_{\zeta(i)}(x, \zeta)] =$$

$$\sum_{i=1}^n \sum_{j \in W} (f(x+j) - f(x + B_i f(\zeta - \zeta_r) + j))^2 + 2nN_w \sigma_n^2$$

În continuare utilizând noile funcții de evaluare vom arăta cum ambiguitățile în stabilirea corespondențelor pot fi eliminate.

În cazul $f(x+j) = f(x+j+a)$, $j \in W$, pentru două sisteme stereo având baseline-ul B_1 , respectiv B_2 , funcția $E[e_{\zeta(12)}(x, \zeta)]$ este:

$$E[e_{\zeta(12)}(x, \zeta)] = \sum_{j \in W} (f(x+j) - f(x + B_1 f(\zeta - \zeta_r) + j))^2 +$$

$$\sum_{j \in W} (f(x+j) - f(x + B_2 f(\zeta - \zeta_r) + j))^2 + 4N_w \sigma_n^2$$

Okutomi și Kanade demonstrează⁴⁷ că:

$$E[e_{\zeta(12)}(x, \zeta)] > 4N_w \sigma_n^2 = E[e_{\zeta(12)}(x, \zeta_r)] \text{ pentru orice } \zeta \neq \zeta_r.$$

O altă ambiguitate va fi rezolvată în felul următor. Considerând funcția $f(x)$ periodică, având perioada T , funcția $e_{\zeta(i)}(x, \zeta)$ va fi la rândul ei periodică, de

perioadă $\frac{T}{B_i f}$. Ceea ce înseamnă că $e_{\zeta(i)}(x, \zeta)$ va avea minime la intervale de

lungime $\frac{T}{B_i f}$. Utilizând, ca mai sus, două sisteme stereo funcția $e_{\zeta(12)}(x, \zeta)$ va

rămâne periodică dar având perioada cel mai mic multiplu comun al perioadelor

$\frac{T}{B_1 f}$ și $\frac{T}{B_2 f}$. Alegând B_1 și B_2 de valori apropiate ambiguitatea este înlăturată.

3.3. Urmărirea trăsăturilor

Odată extrase caracteristicile din fiecare imagine a unei perechi de imagini consecutive I_1, I_2 urmează stabilirea corespondențelor dintre acestea. Se folosește criteriul sumei normalizate a pătratelor diferențelor aplicat pe vecinătăți $M \times M$ ale caracteristicilor a căror corespondență urmează a fi testată:

⁴⁷ Masatoshi Okutomi, Takeo Kanade, „A multiple-baseline stereo” în *IEEE Transactions on Pattern Analysis and Machine Intelligence* (PAMI), vol. 15(4), 1993, pp. 353-363.

$$C(I_1, I_2) = \frac{\sum_{u=-\frac{M}{2}}^{\frac{M}{2}} \sum_{v=-\frac{M}{2}}^{\frac{M}{2}} \left((I_1(u, v) - \bar{I}_1) - (I_2(u, v) - \bar{I}_2) \right)^2}{\sqrt{\sum_{u=-\frac{M}{2}}^{\frac{M}{2}} \sum_{v=-\frac{M}{2}}^{\frac{M}{2}} (I_1(u, v) - \bar{I}_1)^2 \sum_{u=-\frac{M}{2}}^{\frac{M}{2}} \sum_{v=-\frac{M}{2}}^{\frac{M}{2}} (I_2(u, v) - \bar{I}_2)^2}} \quad (118)$$

unde \bar{I}_1 și \bar{I}_2 sunt mediile nivelelor de gri din cele două vecinătăți $M \times M$. De obicei $M=13$.

Se calculează criteriul (118) utilizând candidați pereche pentru colțul curent din prima imagine dintr-o fereastră de căutare de dimensiune 70×70 în cea de a doua imagine centrată în jurul pixelului de aceleași coordonate cu colțul curent din prima imagine. Pentru eliminarea zgomotelor stabilirea perechilor se face aplicând criteriul (118) atât pornind de la imaginea precedentă și cautând într-o fereastră 70×70 în jurul poziției caracteristicii curente în imaginea curentă cât și inversând ordinea celor două imagini. Se aleg corespondențele care au valori maxime locale aproximativ egale în ambele sensuri. Odată stabilite corespondențele se determină, cum se va vedea mai jos, transformata euclidiană dintre sistemele robotului în poziția curentă și poziția precedentă. Utilizând această primă estimare a transformatei euclidiene caracteristicile 3D din baza de date sunt proiectate apoi în imaginea curentă. Se stabilesc iarăși perechile de corespondențe de data această însă pentru ferestre de căutare de 4-5 pixeli și se calculează din nou transformata euclidiană care ar suprapune corespondențele în imagini cât mai exact. Transformata euclidiană finală în pasul respectiv rezultă din corecția primei transformate folosind-o pe cea calculată în pasul doi.

3.4. Estimarea mișcării: calculul poziției robotului în sistemul de referință extern

Estimarea deplasării robotului este un proces iterativ care calculează transformarea euclidiană (rotații și translații) a sistemului robotului în poziția curentă în raport cu sistemul robotului în poziția precedentă. Vectorul de mișcare al robotului, care trebuie determinat, este: $(D_x, D_y, D_z, \alpha_1, \alpha_2, \alpha_3)$, unde (D_x, D_y, D_z) este vectorul translației originii sistemului robotului în raport cu sistemul robotului în poziția precedentă iar $\alpha_1, \alpha_2, \alpha_3$ sunt unghiurile de rotație ale sistemului robotului în noua poziție față de axele OX, OY și OZ ale sistemului robotului în poziția precedentă.

Se pleacă de la o transformată euclidiană inițială. Fie $P_{i+1}(X_{i+1}, Y_{i+1}, Z_{i+1})$ o caracteristică 3D rezultată în pasul 2, exprimată în sistemul robotului din poziția precedentă. Aceeași caracteristică are coordonatele $P_i(X_i, Y_i, Z_i)$ în urma calculului similar din pasul 2 efectuat pentru poziția precedentă a robotului.

În continuare detaliem procedura de calcul a transformatei euclidiene (\mathbf{R} , \mathbf{t}), unde \mathbf{R} este matricea de rotație iar \mathbf{t} vectorul translațiilor.

Avem:

$$\mathbf{P}_{i+1} = \mathbf{R}\mathbf{P}_i + \mathbf{t} \quad (119)$$

Se lucrează în ipoteza unui model general al camerei, modelul proiecției centrale (pinhole camera), unde relațiile dintre coordonatele 3D ale punctelor din lumea reală, exprimate în sistemul de referință al camerei, și coordonatele imaginilor acestora, din planul imaginii, sunt:

$$[x_i, y_i]^t = \left[\frac{fX_i}{Z_i}, \frac{fY_i}{Z_i} \right]^t \quad (120)$$

unde x_i, y_i sunt coordonatele din imagine ale punctelor $P_i(X_i, Y_i, Z_i)$, exprimate în sistemul de referință al camerei iar f este distanța focală, presupusă cunoscută întru-cât camera a fost în prealabil calibrată.

Din relațiile (119) și (120) rezultă:

$$x_{i+1} = \frac{r_{11}X_i + r_{12}Y_i + r_{13}Z_i + t_1}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + t_3}, \quad y_{i+1} = \frac{r_{21}X_i + r_{22}Y_i + r_{23}Z_i + t_2}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + t_3} \quad (121)$$

r_{ij} cu $i, j = 1, 3$ reprezintă elementele matricei \mathbf{R} , exprimate în funcție de unghiurile de rotație $\alpha_1, \alpha_2, \alpha_3$ în jurul celor trei axe ale sistemului de coordonate al robotului în poziția curentă:

$$\begin{aligned} r_{11} &= \cos(\alpha_2)\cos(\alpha_3) \\ r_{12} &= \cos(\alpha_3)\sin(\alpha_1)\sin(\alpha_2) - \cos(\alpha_1)\sin(\alpha_3) \\ r_{13} &= \sin(\alpha_1)\sin(\alpha_3) + \cos(\alpha_3)\cos(\alpha_1)\sin(\alpha_2) \\ r_{21} &= \cos(\alpha_2)\sin(\alpha_3) \\ r_{22} &= \sin(\alpha_1)\sin(\alpha_2)\sin(\alpha_3) + \cos(\alpha_1)\cos(\alpha_3) \\ r_{23} &= \cos(\alpha_1)\sin(\alpha_2)\sin(\alpha_3) - \sin(\alpha_1)\cos(\alpha_3) \\ r_{31} &= -\sin(\alpha_2) \\ r_{32} &= \cos(\alpha_2)\sin(\alpha_1) \\ r_{33} &= \cos(\alpha_1)\cos(\alpha_2) \end{aligned} \quad (122)$$

Problema care se pune este aceea a determinării unghiurilor de rotație $\alpha_1, \alpha_2, \alpha_3$ și a vectorului translațiilor $\mathbf{t} = [t_1, t_2, t_3]^t$, care introduse în (119) minimizează erorile:

$$\begin{aligned} \delta x_i &= x_i(\alpha_1, \alpha_2, \alpha_3, t_1, t_2, t_3) - x_{i+1} \\ \delta y_i &= y_i(\alpha_1, \alpha_2, \alpha_3, t_1, t_2, t_3) - y_{i+1} \end{aligned} \quad (123)$$

În (123) x_{i+1}, y_{i+1} sunt imaginile trăsăturilor extrase în poziția curentă, corespunzătoare trăsăturilor x_i, y_i din poziția precedentă.

Pornind de la o primă aproximare sau soluție inițială a matricei de rotații \mathbf{R}^0 și a vectorului translațiilor \mathbf{t}^0 erorile (123) pot fi approximate printr-o dezvoltare Taylor de ordinul întâi într-o vecinătate a parametrilor curenți așa încât:

$$\sum_{j=1}^3 \left[\frac{\partial x_i}{\partial t_j} \Delta t_j + \frac{\partial x_i}{\partial \alpha_j} \Delta \alpha_j \right] = \delta x_i, \quad \sum_{j=1}^3 \left[\frac{\partial y_i}{\partial t_j} \Delta t_j + \frac{\partial y_i}{\partial \alpha_j} \Delta \alpha_j \right] = \delta y_i \quad (124)$$

Se urmărește așadar determinarea variațiilor $\Delta \alpha_j, \Delta t_j, j=1,3$ din (124) în condițiile în care $\delta x_i, \delta y_i$ se cunosc din (123), pentru valorile curențe ale unghiurilor de rotație, respectiv translațiilor, iar derivatele $\frac{\partial x_i}{\partial t_j}, \frac{\partial x_i}{\partial \alpha_j}, \frac{\partial y_i}{\partial t_j}, \frac{\partial y_i}{\partial \alpha_j}$ au o

expresie analitică cunoscută ori pot fi approximate prin diferențe finite pentru a fi calculate sau estimate pentru valorile curențe ale parametrilor. După calculul variațiilor $\Delta \alpha_j, \Delta t_j$, noile valori ale translațiilor se calculează după formula:

$t_{j+1} = t_j + \Delta t_j$ iar cu valorile $\Delta \alpha_j$ se formează o matrice provizorie a rotațiilor care înmulțită la dreapta cu \mathbf{R}_j , matricea curentă a rotațiilor, va conduce la o nouă valoare \mathbf{R}_{j+1} a rotațiilor. Procesul este reluat până când eroarea (123) scade sub un anumit prag sau se depășește un număr maxim de iterații, stabilit anterior. Pe parcursul procesului iterativ sunt reținute valorile α_j și t_j care au condus la cele mai mici valori ale erorilor (123). Metoda iterativă de mai sus este metoda clasică de optimizare Gauss-Newton.

Lowe propune⁴⁸ o altă metodă de rezolvare a ecuației (119), în forma:

$$[X_{i+1}, Y_{i+1}, Z_{i+1}]^t = \mathbf{R}(\mathbf{P}_i - \mathbf{t}) \quad (125)$$

Translațiile de data aceasta sunt efectuate înaintea rotațiilor. Ecuația de mai sus este reparametrizată în sensul că $(\mathbf{P}_i - \mathbf{t})$ devine \mathbf{P}_i iar componentele \mathbf{t} sunt exprimate în sistemul de coordonate al imaginii și nu al sistemului robotului. Deci (125) devine:

$$[X_{i+1}, Y_{i+1}, Z_{i+1}]^t = \mathbf{R}\mathbf{P}_i \quad (126)$$

iar (120) se transformă în:

$$[x_{i+1}, y_{i+1}]^t = \left[\frac{fX_i}{Z_i + D_z} + D_x, \frac{fY_i}{Z_i + D_z} + D_y \right]^t \quad (127)$$

Prin urmare se pune problema determinării translațiilor D_x, D_y, D_z unghiurile fiind aceleași ca și în cazul precedent, adică $\alpha_1, \alpha_2, \alpha_3$. De notat că translația pe axa Z își păstrează semnificația din cazul clasic.

Relația dintre \mathbf{t} din (125) și D_x, D_y, D_z este dată de:

⁴⁸ David G. Lowe, „Fitting Parameterized Three-Dimensional Models to Images” în *IEEE Transactions on Pattern Analysis and Machine Intelligence* (TPAMI) vol. 13(5), 1991, pp. 441-450.

$$\mathbf{T} = \mathbf{R}^{-1} \left[-\frac{D_x(Z + D_z)}{f}, -\frac{D_y(Z + D_z)}{f}, -D_z \right] \quad (128)$$

Avantajul acestei parametrizări constă în faptul că derivatele coordonatelor X_i, Y_i, Z_i deci și ale imaginilor x_i, y_i sunt extrem de simple în raport cu unghiurile de rotație. De exemplu, derivata lui X_i în raport cu α_3 , unghiul de rotație în jurul axei Z , este $-Y_i$. Aceasta rezultă din faptul că:

$$(X_i, Y_i, Z_i) = (r \cos \alpha_3, r \sin \alpha_3, Z_i),$$

unde r este distanța punctului față de axa Z .

$$\text{Deci } \frac{\partial X_i}{\partial \alpha_3} = -r \sin \alpha_3 = -Y_i.$$

În tabelul de mai jos sunt date toate derivatele coordonatelor X_i, Y_i, Z_i în raport cu unghiurile de rotație.

	X_i	Y_i	Z_i
α_1	0	$-Z_i$	Y_i
α_2	Z_i	0	$-X_i$
α_3	$-Y_i$	X_i	0

Tabelul 1. Derivatele coordonatelor X_i, Y_i, Z_i în raport cu unghiurile de rotație $\alpha_1, \alpha_2, \alpha_3$.

Calculul derivatelor coordonatelor din imagine ale punctelor este destul de simplu. De exemplu, din (127) rezultă:

$$x_i = \frac{fX_i}{Z_i + D_z} + D_x \Rightarrow \frac{\partial x_i}{\partial D_x} = 1$$

$$\frac{\partial x_i}{\partial \alpha_2} = \frac{f}{Z_i + D_z} \frac{\partial X_i}{\partial \alpha_2} - \frac{fX_i}{(Z_i + D_z)^2} \frac{\partial Z_i}{\partial \alpha_2}$$

$$\text{Din tabelul 1: } \frac{\partial X_i}{\partial \alpha_2} = Z_i \text{ și } \frac{\partial Z_i}{\partial \alpha_2} = -X_i$$

$$\text{Deci: } \frac{\partial x_i}{\partial \alpha_2} = f \left(\frac{Z_i}{Z_i + D_z} - \frac{X_i^2}{(Z_i + D_z)^2} \right), \text{ unde } c = \frac{1}{Z_i + D_z}$$

În tabelul următor sunt figurate derivatele parțiale ale coordonatelor din imagine x_i și y_i în funcție de parametri de calculat, $D_x, D_y, D_z, \alpha_1, \alpha_2, \alpha_3$. Din acestea se construiește iacobianul. Revenirea la formula (125) se face aplicând (127).

	x_i	y_i
D_x	1	0
D_y	0	1
D_z	$-fc^2 X_i$	$-fc^2 Y_i$
α_1	$-fc^2 X_i Y_i$	$-fc(Z_i + cY_i^2)$
α_2	$fc(Z_i + cX_i^2)$	$fc^2 X_i Y_i$
α_3	$-fc Y_i$	$fc X_i$

Tabelul 2. Derivatele parțiale ale coordonatelor din imagine în funcție de translații și unghiurile de rotație

Așa cum s-a spus mai sus folosind această primă estimare a transformatei euclidiene caracteristicile 3D din baza de date sunt proiectate apoi în imaginea curentă. Se stabilesc iarăși perechile de corespondențe de data această însă pentru ferestre de căutare reduse, de 4-5 pixeli și se calculează din nou transformata euclidiană care ar suprapune corespondențele în imagini cât mai exact. Transformata euclidiană finală în pasul respectiv rezultă din corecția primei transformate folosind-o pe cea calculată în pasul doi.

În etapa a doua, de rafinare a transformatei euclidiene care indică poziția robotului în poziția curentă în raport cu poziția precedentă se memorează erorile dintre proiecțiile în imaginea curentă a coordonatelor 3D ale trăsăturilor și corespondențele lor din imaginea curentă adică:

$$E_x = x_{i+1} - \left(\frac{fX_i}{Z_i + D_z} + D_x \right) \quad E_y = y_{i+1} - \frac{fY_i}{Z_i + D_z} + D_y$$

Începând cu a patra iterație sunt eliminate perechile pentru care eroarea $\sqrt{E_x^2 + E_y^2}$ este semnificativă și procesul continuă cu restul corespondențelor.

La sfârșitul procesului iterativ poziția robotului este memorată în baza de date. Totodată se reține și poziția trăsăturilor din imaginea curentă. Baza de date este de asemenea actualizată în sensul incrementării contorului de vizibilitate ale tuturor trăsăturilor vizibile în imaginea curentă, decrementării contorului trăsăturilor a căror proiecție este în afara imaginii curente și eliminării trăsăturilor invizibile în ultimele cinci poziții ale robotului.

3.1. Corecția poziției robotului și a coordonatelor 3D ale trăsăturilor scenei

Pentru a contracara erorile de măsurare ale caracteristicilor scenei și de localizare a robotului, deci de trasare a traiectoriei lui, se folosesc filtre Kalman. Fiecărei trăsături active din baza de date i se asociază un filtru Kalman. De asemenea, un alt filtru Kalman este folosit pentru corecția poziției robotului. Filtrul Kalman folosit pentru rafinarea poziției sistemului stereo (robotului) este:

$$\mathbf{X}_{k+1} = \mathbf{F}\mathbf{X}_k + \xi_k \quad (129)$$

$$\mathbf{z}_k = \mathbf{H}\mathbf{X}_k + \eta_k \quad (130)$$

unde $\mathbf{X}_k = (X, Y, Z, \alpha_1, \alpha_2, \alpha_3, \dot{X}, \dot{Y}, \dot{Z}, \dot{\alpha}_1, \dot{\alpha}_2, \dot{\alpha}_3)$ este variabila de stare în poziția k a robotului, \mathbf{F} este o matrice constantă 12×12 definită de Saeedi, Lawrence și Lowe⁴⁹, ξ_k și η_k sunt vectorii zgomotelor aferente sistemului și observațiilor, \mathbf{H} este o matrice 6×12 de forma $[\mathbf{I} \ \mathbf{N}]$

cu \mathbf{I} matrice unitate 6×6 și \mathbf{N} matrice nulă 6×6 .

Predicția filtrului este:

$$\mathbf{X}(k+1 | k) = \mathbf{F}\mathbf{X}(k | k) \quad (131)$$

Notând cu $\mathbf{P}(k | k)$ matricea de covarianță a procesului covarianța predicției este:

$$\mathbf{P}(k+1 | k) = \mathbf{F}\mathbf{P}(k | k)\mathbf{F}^t + \mathbf{Q}(k) \quad (132)$$

unde \mathbf{Q} este o matrice 12×12 constantă determinată experimental, asociată zgomotului procesului.

Predicția măsurătorilor este dată de:

$$\mathbf{z}(k+1 | k) = \mathbf{H}\mathbf{X}(k+1 | k) \quad (133)$$

Noua poziție a robotului, \mathbf{X}_{LS} , se obține actualizând vechea poziție $\mathbf{X}(k | k)$ cu transformata dată de $D_x, D_y, D_z, \alpha_1, \alpha_2, \alpha_3$ calculați în etapa 6 de estimare a mișcării robotului. Notând transformarea respectivă cu $(\mathbf{D}_x, \mathbf{R}_x)$ avem:

$$\mathbf{X}_{LS} = [\mathbf{R}_x][\mathbf{X}(k | k)] + \mathbf{D}_x \quad (134)$$

Matricea de covarianță a măsurătorilor, notată \mathbf{R}_{LS} , se actualizează în etapa a doua a procesului de determinare a poziției robotului.

Notând cu \mathbf{J} matricea sistemului (124) avem $\mathbf{R}_{LS} = (\mathbf{J}^t \mathbf{J})^{-1}$.

Pentru fiecare poziție a robotului filtrul este actualizat după cum urmează:

⁴⁹ Parvaneh Saeedi, Peter D. Lawrence, David G. Lowe, „Vision-based 3-D trajectory tracking for unknown environments” în *IEEE Transactions on Robotics*, feb. 2006, vol. 22, pp. 119-136.

$$\mathbf{P}(0 | 0) = \text{Var}(\mathbf{X}_0)$$

$$\mathbf{P}(k + 1 | k) = \mathbf{F}\mathbf{P}(k | k)\mathbf{F}^t + \mathbf{Q}(k)$$

$$\mathbf{W}(k + 1) = \mathbf{P}(k + 1 | k)\mathbf{H}^t [\mathbf{H}\mathbf{P}(k + 1 | k)\mathbf{H}^t + \mathbf{R}_{LS}]^{-1}$$

$$\mathbf{P}(k + 1 | k + 1) = \mathbf{P}(k + 1 | k) - \mathbf{W}(k + 1)\mathbf{H}\mathbf{P}(k + 1 | k)$$

$$\mathbf{X}(k + 1 | k) = \mathbf{F}\mathbf{X}(k | k)$$

$$\mathbf{z}(k + 1 | k) = \mathbf{H}\mathbf{X}(k + 1 | k)$$

$$\mathbf{X}(k + 1 | k + 1) = \mathbf{X}(k + 1 | k) + \mathbf{W}(k + 1)(\mathbf{X}_{LS} - \mathbf{z}(k + 1 | k))$$

Aşa cum s-a spus câte un filtru Kalman este asociat şi fiecărei trăsături statice din scena reală, trăsătură activă în procesul stabilirii traiectoriei adică vizibilă în imaginea curentă. Filtrul asociat trăsăturii $\mathbf{w}(X, Y, Z)$ este descris de ecuaţiile:

$$\mathbf{C}_{k+1}^{-1} = \mathbf{C}_k^{-1} + \mathbf{A}_{k+1}^t \mathbf{V}_{k+1}^{-1} \mathbf{A}_{k+1}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mathbf{k}_{k+1} (b_{k+1} - \mathbf{A}_{k+1} \mathbf{w}_k)$$

$$\mathbf{k}_{k+1} = \mathbf{C}_{k+1} \mathbf{A}_{k+1}^t \mathbf{V}_{k+1}$$

unde \mathbf{C}_k exprimă incertitudinea estimării coordonatelor trăsăturii $\mathbf{w}(X, Y, Z)$, \mathbf{k}_k este câştigul filtrului, b_k reprezintă coordonatele 3D rezultate din măsurătorile din imagine iar \mathbf{V}_k este covarianţa erorilor.

Concluzii

Capitolul este o sinteză a principalelor tehnici de reconstrucţie din imagini multiple ale unor obiecte în mişcare, tehnici prezentate în literatura de specialitate. Sunt discutate mai întâi cele mai performante tehnici de reconstrucţie afină din imagini obţinute în principal prin proiecţie ortografică. Se porneşte de la analiza mişcării unui singur corp rigid, după care studiul este extins la analiza mişcării unui număr necunoscut de obiecte rigide. În ceea ce priveşte reconstrucţia proiectivă studiul se rezumă la analiza mişcării unui singur corp rigid. În ultima parte este prezentată o aplicaţie de urmărire a unui robot mobil, pornind de la lucrarea „Mobile Robot Localization and Mapping with Uncertainty using Scale-Invariant Visual Landmarks”⁵⁰.

⁵⁰ Stephen Se, David G. Lowe, James J. Little, „Mobile Robot Localization and Mapping with Uncertainty using Scale-Invariant Visual Landmarks” în International Journal of Robotic Research, vol. 21(8), 2002, pp. 735-760.

Bibliografie

- Christian Banz, Sebastian Hesselbarth, Holger Flatt, Holger Blume, Peter Pirsch, „Real-time stereo vision system using semi-global matching disparity estimation: Architecture and FPGA-implementation” în *2010 International Conference on Embedded Computer Systems (SAMOS)*, nov. 2010, pp. 93-101.
- Richard Berthilsson, Anders Heyden, Gunnar Sparr, „Recursive Structure and Motion from Image Sequences using Shape and Depth Spaces” în *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 1997, pp. 444-449.
- Joao Costeira, Takeo Kanade, „A Multi-body Factorization Method for Motion Analysis” în *International Journal of Computer Vision*, 29(3), 1998, pp. 159-179.
- David A. Forsyth, Jean Ponce, „Computer Vision. A modern approach”, *Prentice Hall*, 2003.
- C. William Gear, „Feature grouping in moving objects” în *Proceedings of the 1994 IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, 1994, pp. 214-219.
- Chris Harris, Mike Stephens, „A combined corner and edge detector” în *Proceedings of 4th Alvey Vision Conference*, 1988, pp. 147-151.
- Richard Hartley, Andrew Zisserman, „Multiple View Geometry in Computer Vision”, Cambridge University Press, 2004, Second Edition.
- Anders Heyden, „Reconstruction from multiple images by means of using relative depths” în *International Journal of Computer Vision*, 24(2), 1997, pp. 155-161.
- Jae-Hak Kim, Lourdes Agapito, „Motion Segmentation using the Hadamard Product and Spectral Clustering” în *Proceedings of the 2009 international conference on Motion and video computing (WMVC 2009)*, December 2009, Snowbird, Utah, USA, pp. 126-133.
- Jan J. Koendering, Andrea J. van Doorn, „Affine Structure from Motion” în *Journal of the Optical Society of America A*, 8, pp. 377-385, 1991.
- Fabien Lauer, Christoph Schnorr, „Spectral clustering of linear subspaces for motion segmentation” în *The 12th IEEE International Conference on Computer Vision (ICCV'09)*, 2009, pp. 678-675.
- David G. Lowe, „Fitting Parameterized Three-Dimensional Models to Images” în *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* vol. 13(5), 1991, pp. 441-450.
- Shzjan Mahamud, Martial Herbert, „Iterative projective reconstruction from multiple views” în *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00)*, June, 2000, pp. 430- 437.
- Don Murray, Jim Little, „Using Real-Time Stereo Vision for Mobile Robot Navigation” în *Journal of Autonomous Robots*, Vol.8, Issue 2, April 2000, pp. 161-171.
- Andrew Y. Ng, Michael I. Jordan, Yair Weiss, „On spectral clustering: Analysis and an algorithm” în *Advances in Neural Information Processing Systems (NIPS'02)*, MIT Press, 2002, vol. 14, pp. 849-856.
- Masatoshi Okutomi, Takeo Kanade, „A multiple-baseline stereo” în *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (PAMI), vol. 15(4) 1993, pp. 353-363.
- Florin Rotaru, Dan Gâlea, Ion Păvăloi, Silviu Bejinariu, Cristina Niță, Ramona Luca, „Reconstrucție 3D din două imagini” în *Raport de cercetare*, noiembrie 2007, IIT Iași.
- Florin Rotaru, Silviu Bejinariu, Cristina Niță, Ramona Luca, Camelia Lazar, Simona Pescaru, „Tehnici de reconstrucție afină și proiectivă din imagini multiple pentru estimarea mișcării și fuziunea informațiilor multi-senzor” în *Raport de cercetare*, iunie 2010, IIT Iași.
- Florin Rotaru, Silviu Bejinariu, Cristina Niță, Ramona Luca, Camelia Lazar, „Tehnici de analiză a mișcării, fuziunea informațiilor, procese fundamentale de interacțiune a undelor, logici pentru reprezentarea cunoștințelor. Aplicații și implementare instrumente software” în *Raport de cercetare*, noiembrie 2011, IIT Iași.
- Parvaneh Saeedi, Peter D. Lawrence, David G. Lowe, „3D Motion Tracking of a Mobile Robot in a Natural Environment” în *Proceedings of the 2000 IEEE International Conference on Robotics and Automation, ICRA 2000*, pp. 1682-168.
- Parvaneh Saeedi, David G. Lowe, Peter D. Lawrence, „An efficient binary corner detector” în *Proceedings of Seventh International Conference on Control, Automation, Robotics and Vision, ICARCV 2002*, pp. 338-343.

FLORIN ROTARU, SILVIU BEJINARIU, SIMONA PESCARU

- Parvaneh Saeedi, David G. Lowe, Peter D. Lawrence, „3D localization and tracking in unknown environments” in *Proceedings of the 2003 IEEE International Conference on Robotics and Automation*, ICRA 2003, pp. 1297-1303.
- Parvaneh Saeedi, Peter D. Lawrence, David G. Lowe, „Vision-based 3-D trajectory tracking for unknown environments” in *IEEE Transactions on Robotics*, Feb. 2006, Vol. 22, pp. 119-136.
- Stephen Se, David G. Lowe, James J. Little, „Mobile Robot Localization and Mapping with Uncertainty using Scale-Invariant Visual Landmarks” in *International Journal of Robotic Research*, vol. 21(8), 2002, pp. 735-760.
- Peter Sturm, Bill Triggs, „A Factorization Based Algorithm for Multi-Image Projective Structure and Motion” in *Proceedings of 4th European Conference on Computer Vision*, Cambridge, England, April 1996, pp. 709-720.
- Satoshi Takezawa, Mudiyansele Herath, Gamini Dissanayake, „SLAM in indoor environments with stereo vision” in *Proceedings of 2004 International Conference on Intelligent Robots and Systems*, (IROS 2004), vol. 2, pp. 1866-1871.
- Celine Teuliere, Lucile Martin, Christophe Leroux, Edvige Pissaloux, „Vision and Inertial Sensor Fusion for 3D Self-Localization in Unknown Environment” in *9th ESA Workshop on Advanced Space Technologies for Robotics and Automation*, ASTRA 2006, Noordwijk, The Netherlands, November 2006.
- Carlo Tomasi, Takeo Kanade, „Shape and motion from image streams under orthography: A factorization approach” in *International Journal of Computer Vision*, Vol 9, Issue 2, November 1992, pp. 137-154.

Metode de analiză și extragere a caracteristicilor esențiale din imaginile digitale

ADRIAN CIOBANU¹

ABSTRACT. Due to the availability of digital cameras there is a tremendous increase in the number of available digital images, for instance, on the Internet. Finding a particular image can be almost impossible by browsing, as searching through text is not available for the majority of the images. There is no time for manually annotate all the available images, so the only solution is the automatic annotation based on image content, i.e. based on color, texture, shape and other features. These features must be extracted for each image and stored in a database along with the images. As the number of images is so big, the features must be very economic in computational time and space required to be stored. In this chapter a review of the existing feature extraction methods that qualify as economic is presented as a necessary step before trying to propose new methods in this field.

KEYWORDS: *feature extraction, color features, texture features, shape features, image retrieval, similarity.*

1. Introducere

Analiza și procesarea imaginilor constituie unul dintre cele mai interesante și dinamice domenii din cadrul științei calculatoarelor. Interesul pentru acest domeniu provine fie din nevoia de a îmbunătăți informația vizuală destinată publicului, fie din nevoia de a procesa imaginile în vederea stocării, transmiterii și reprezentării lor pentru uzul unor echipamente avansate, cum ar fi roboții industriali, vehiculele autonome etc. [Gonzalez, 2002]². În ultimele decenii se constată o maturizare a domeniului procesării de imagini prin implementarea unui număr din ce în ce mai mare de metode specifice acestuia în cadrul aplicațiilor software utilizate frecvent, cum ar fi: teledetecția, diagnoza industrială, controlul vehiculelor, imagistica biomedicală și supravegherea automată, pentru a enumera aici doar câteva dintre subdomeniile mai importante. Progresul în procesarea imaginilor se poate măsura și prin numărul din ce în ce mai mare de programe de calculator și echipamente specifice disponibile pe piață. Exemplul cel mai edificator este actuala avalanșă de aparate foto digitale, modele comerciale sau profesionale, cu caracteristici

¹ Adrian Ciobanu, cercetător științific, IIT, Academia Română, filiala Iași.

² Gonzalez, Rafael C., Woods, Richard E., „*Digital Image Processing*”, Prentice Hall, Upper Saddle River, NJ, 2002.

funcționale din ce în ce mai avansate, însoțite de programele de calculator aferente, destinate prelucrării și tipăririi imaginilor digitale achiziționate [Sonka, 2008]³. În spatele acestei dezvoltări a domeniului stau însă eforturi susținute, mai ales de ordin experimental, necesare pentru a găsi cele mai viabile soluții la problemele ce se cer rezolvate prin procesare de imagini [Gonzalez, 2009]⁴.

Una dintre primele utilizări ale imaginilor digitale datează încă din anii 1920 (cu mult înainte de apariția calculatoarelor!) și a constat în transmiterea de fotografii referitoare la evenimentele sportive de la Londra la New York, prin cablul submarin care traversa Atlanticul, cu ajutorul sistemului Bartlane. Se reducea durata transportului unei fotografii de la o săptămână la mai puțin de 3 ore. Imaginile erau codate cu echipamente speciale la un capăt al cablului submarin, după care la celălalt capăt erau decodate și tipărite cu o imprimantă telegrafică în doar 5 tonuri de gri (tehnologie care simula „halftone pattern”-ul utilizat mult timp după aceea la tipărirea fotografiilor din ziare). Până în 1929 tehnologia a evoluat, ajungându-se până la 15 tonuri de gri.

Bazele calculatoarelor moderne au fost puse în 1940 de John von Neumann (prin conceptul de memorie pentru program și date și prin conceptul de salt condițional), dar primele calculatoare care au fost suficient de puternice pentru a prelucra imagini au apărut de abia în anii '60. Utilizarea calculatoarelor în acest domeniu s-a datorat cercetărilor din domeniul spațial, efectuate la Jet Propulsion Laboratory (Pasadena, California), unde au fost corectate, cu ajutorul calculatorului, primele imagini transmise de pe Lună de modulul Ranger 7, la data de 31 iulie 1964. Deja la începutul anilor 1970 au apărut și primele aplicații de imagistică medicală, în special cele legate de tomografia computerizată (CT).

Din acel moment aplicațiile procesărilor de imagini s-au diversificat din ce în ce mai mult: aplicații medicale și industriale ale razelor X, imagistica aeriană și satelitară, restaurarea fotografiilor în arheologie, imagistica experimentelor cu plasmă de înaltă energie, microscopia electronică etc. Ele au devenit uzuale în astronomie, medicină nucleară, apărare și industrie. S-au dezvoltat și aplicații de prelucrare de imagini specifice funcționării unor echipamente speciale, pe ramura denumită în engleză „machine vision” sau „machine perception”: recunoașterea automată a caracterelor (OCR), aplicații pentru roboții de pe liniile de asamblare și control industrial, recunoaștere militară, procesarea automată a amprentelor digitale, evaluarea automată a radiografiilor și probelor de sânge, procesarea automată a imaginilor satelitare pentru prezicerea vremii sau pentru evaluarea mediului înconjurător [Gonzalez, 2002]⁵.

³ Sonka, Milan, Hlavac, Vaclav, Boyle, Roger, „*Image Processing, Analysis, and Machine Vision*”, International Student Edition, Thomson, 2008.

⁴ Gonzalez, Rafael C., Woods, Richard E., and Eddins, Steven L., „*Digital Image Processing using MATLAB*”, Gatesmark Publishing, 2009.

⁵ Gonzalez, Rafael C., Woods, Richard E., „*Digital Image Processing*”, Prentice Hall, Upper Saddle River, NJ, 2002.

1.2 Imagini digitale

Pentru a simplifica introducerea, considerăm mai întâi o imagine digitală doar în tonuri de gri. O imagine poate fi definită ca o funcție bidimensională $f(x,y)$, x și y fiind coordonatele unui punct din cadrul imaginii și mărimea $f(x,y)$ fiind denumită intensitatea sau nivelul de gri al imaginii în punctul determinat de perechea (x,y) . Dacă valorile x , y , și $f(x,y)$ sunt finite sau discrete, atunci se poate spune despre o imagine că este digitală. Practic, o imagine digitală este formată dintr-un număr finit de elemente care au fiecare o anumită locație și intensitate, denumite elemente de imagine sau pixeli. Termenul pixel (derivat din englezescul „picture element”) este cel mai încetățenit la ora actuală și va fi folosit în acest capitol.

Vederea este cel mai avansat simț al omului și de aceea nu este întâmplător rolul extrem de important pe care îl joacă imaginile în percepția noastră despre lume [Gonzalez, 2002]⁶. Oamenii sunt capabili de a sesiza o anumită bandă din spectrul electromagnetic al radiației luminoase, de aceea o importanță deosebită o au imaginile color. În acest caz funcția $f(x,y)$ nu mai are o singură valoare, ci poate avea trei valori, sub forma unui triplet (R,G,B) , unde fiecare componentă reprezintă nivelul de gri din cadrul spectrului roșu, verde și, respectiv, albastru. Este cazul spațiului de culoare RGB, unul dintre cele mai utilizate în afișarea imaginilor digitale color pe ecranele de televiziune sau pe calculatoarele uzuale.

Însă, în cazul unor echipamente speciale, imaginile nu sunt limitate la spectrul vizual uman, acoperind întregul spectru electromagnetic, de la razele gamma și până la undele radio. În plus, au apărut numeroase surse de imagini neconvenționale generate în cadrul ultrasonografiei, microscopiei electronice sau chiar generate pur și simplu de calculator.

1.3 Problematika regăsirii imaginilor digitale

Utilizarea imaginilor în cadrul comunicării interumane datează încă din vremuri preistorice, exemplul clasic fiind picturile rupestre. Chiar înainte de epoca romană au fost folosite hărți și planuri de construcție a imobilelor. Totuși, niciodată nu a existat o creștere atât de importantă a numărului de imagini digitale, a gradului lor de accesibilitate și a importanței acestora în toate aspectele vieții de zi cu zi, cum s-a întâmplat în ultimii zece ani. Dacă inițial motorul acestei evoluții a fost fotografia și televiziunea, rolul a fost preluat apoi de calculatoarele personale și dus la extrem de multitudinea de aparate fotografice digitale disponibile acum. Odată cu dezvoltarea Internetului, s-a ajuns ca un utilizator obișnuit să poată accesa acum miliarde de fotografii pe un simplu laptop. O primă aplicație de prelucrare a imaginilor (Sketchpad – aplicație de creare, manipulare și înmagazinare a imaginilor, dezvoltată de Ivan Sutherland [Eakins, 1999]⁷) a apărut încă din 1965, însă prețul exorbitant al echipamentelor necesare a făcut ca de abia pe la mijlocul anilor 1980 să înceapă să

⁶ *Idem.*

⁷ Eakins, John, Graham, Margaret, „JISC Technology Applications Programme Report #39”, University of Northumbria at Newcastle, October 1999.

fie utilizate astfel de aplicații pe scară largă. Valoarea echipamentelor a devenit accesibilă doar datorită industriei jocurilor de calculator, care a dus la comercializarea în masă a calculatoarelor și la scăderea prețurilor acestora, inclusiv a componentei celei mai importante a acestora, placa video.

Faptul că imaginile sunt acum cu preponderență de tip digital nu înseamnă că sunt mai ușor de gestionat. Din contră, rata mare de creștere a numărului de fotografii digitale disponibile face improprie căutarea sau regăsirea lor prin răsfoire (browsing), iar metoda indexării lor pe bază unor adnotări textuale nu poate face nici ea față în contextul exploziei numărului de imagini disponibile pe Internet, deși este singura disponibilă. Pur și simplu nu mai este timp pentru ca un operator uman să adnoteze, printr-un număr oarecare de cuvinte cheie, fiecare imagine digitală existentă. De aceea, o soluție la această problemă este considerată autoindexarea imaginilor digitale pe baza conținutului lor, adică pe baza caracteristicilor de culoare, textură și formă ce pot fi extrase automat dintr-o imagine digitală.

Există însă mai multe tipuri de attribute ale imaginilor digitale care ar putea fi folosite în procesul regăsirii lor ca date de intrare, cum ar fi: prezența unei anumite combinații de culori, texturi și forme (de exemplu, benzi orizontale galbene), prezența unui anumit aranjament de obiecte (scaune în jurul unei mese), apartenența la un anumit tip de evenimente (în cadrul unui meci de fotbal), prezența unei personalități, o anumită locație sau un anumit eveniment (prezența lui Ilie Năstase la BCR Open), emoții subiective asociate cu o imagine (fericire) și metainformațiile de tipul datei la care a fost făcută fotografia sau imaginea, de către cine și în ce loc (exprimat, eventual în coordonate geostaționare).

În mod similar, există mai multe tipuri de căutări, clasificate pe trei nivele [Eakins, 1999]⁸: nivelul 1, al caracteristicilor primitive, de gen culoare, textură, formă și al interrelațiilor spațiale între elementele din imagine; nivelul 2, care constă în caracteristici derivate sau logice privitoare la identitatea obiectelor conținute de imaginile digitale; și nivelul 3, care constă în caracteristici abstracte, care implică un raționament de nivel înalt cu privire la înțelesul și rolul pe care îl joacă obiectele în cadrul scenelor reprezentate de imaginile digitale. În numeroase lucrări, nivelele 2 și 3 sunt referite ca regăsire semantică a imaginilor, și este cât se poate de clasică formularea „decalaj semantic” (semantic gap) în ceea ce privește saltul de la nivelul 1 la nivelul 2 de interogare. Decalajul semantic este foarte important și nu a putut fi încă depășit de sistemele de regăsire a imaginilor pe bază de conținut (CBIR – *Content Based Image Retrieval*).

1.4 Trăsături reprezentative pentru imaginile digitale color

Orice sistem de regăsire a imaginilor are la bază un proces de prelucrare a acestora, în vederea extragerii unor caracteristici sau trăsături care să le reprezinte. Acest proces de atribuire a unui set de descriptori de identitate (indecși) pentru fiecare imagine, cu scopul regăsirii unor imagini relevante și al îndepărtării celor

⁸ Eakins, John, Graham, Margaret, „JISC Technology Applications Programme Report #39”, University of Northumbria at Newcastle, October 1999.

irelevante, se numește indexare. Indecșii sunt memorați în baza de date, alături de imagini, și este de preferat să fie proiectați în vederea eficientizării regăsirii imaginilor. Trăsăturile pot fi de origine diferită: de culoare, de formă, de textură, dimensionale, de distanță, de poziție relativă etc. și pot coexista în cadrul bazei de date pentru o descriere cât mai completă a imaginilor.

În momentul în care baza de date este interogată pe baza unei imagini, fie se calculează trăsăturile ei pe baza aceluiași proces de prelucrare aplicat imaginilor din baza de date, în cazul în care avem imaginea este nouă, fie sunt extrase din baza de date trăsăturile ei deja calculate, dacă imaginea a fost deja indexată în baza de date. Sarcina de regăsire a imaginilor asemănătoare este îndeplinită de un motor de evaluare a asemănării imaginilor, care compară trăsăturile imaginii de intrare cu trăsăturile celorlalte imagini din baza de date. Mecanismul de asemănare a imaginilor este implementat pe baza unei anumite metrici sau a unei anumite măsuri de similaritate. Ca rezultat, imaginile din baza de date sunt ordonate în funcție de coeficientul lor de similaritate/potrivire cu imaginea de intrare și un anumit număr dintre imaginile cele mai similare sunt prezentate utilizatorului pentru evaluare.

2. Trăsături de culoare

Similaritatea imaginilor digitale color poate fi evaluată pe baza definirii unor trăsături de culoare robuste și eficiente care să reprezinte conținutul coloristic al imaginilor. Deocamdată nu a fost descoperită cea mai bună reprezentare, recurgându-se la mai multe tipuri de reprezentări care caracterizează culoarea unei imagini din perspective diferite. De regulă însă, orice caracteristică a unei imagini este necesar să aibă următoarele proprietăți de bază:

- similaritate perceptuală – distanța dintre trăsăturile extrase pentru două imagini să fie mare doar dacă imaginile nu sunt similare;
- eficiență – să poată fi calculate într-un timp scurt;
- economie – să aibă dimensiuni mici, pentru a nu fi afectată eficiența regăsirii imaginilor;
- scalabilitate – performanța sistemului să nu depindă foarte mult de numărul de imagini din baza de date;
- robustețe – schimbarea condițiilor de achiziție a imaginilor digitale să nu afecteze regăsirea acestora [Schettini, 2001]⁹.

Trăsăturile de culoare trebuiesc însoțite de o funcție de similaritate/disimilaritate de tip distanță, care să satisfacă următoarele proprietăți pentru oricare trei imagini I, J și K din baza de date:

1. $d(I, I), d(I, J), d(J, J), d(I, K), d(J, K), d(K, K) \geq 0$
2. $d(I, I) = d(J, J) = d(K, K)$ – auto-simetrie

⁹ Schettini, Raimondo, Ciocca, Gianluigi, Zuffi, Silvia, „Indexing and Retrieval in Colour Image Databases”, Chapter 10 in Colour Image Science – Exploiting Digital Media, Editori Lindsay W. MacDonald, M. Ronnier Luo, John Wiley & Sons, England, 2001, pp. 183-213.

3. $d(I, J), d(I, K) \geq d(I, I)$ – minimalitate
4. $d(I, J) = d(J, I), d(I, K) = d(K, I)$ – simetrie
5. $d(I, K) + d(K, J) \geq d(I, J)$ – inegalitate triunghiulară

Proprietățile 1, 2, 3 și 5 definesc o metrică, iar proprietățile 1, 2, 3, 4 definesc o măsură de similaritate [Schettini, 2001]¹⁰.

Indexarea imaginilor pe bază de culoare are la bază faptul că, de cele mai multe ori, culoarea codifică o anumită funcționalitate sau caracterizează un anumit lucru. De exemplu, semnele de circulație folosesc anumite culori, drumurile au o anumită culoare, pădurile sunt verzi, cerul este albastru etc. Deoarece culoarea nu ne permite și identificarea obiectelor, se recurge și la caracteristici de textură și formă pentru a realiza identificarea. Cu toate acestea, culoarea are o importanță primordială în indexarea imaginilor datorită rapidității cu care poate fi analizată, rezultatele căutării coloristice în baza de imagini putând fi ulterior filtrate prin aplicarea unor metode mai lente care implică segmentarea, textura și forma, pentru îmbunătățirea calității rezultatelor [Stricker, 1995]¹¹.

2.1 Reducerea numărului de culori dintr-o imagine

Orice demers de extragere a unor caracteristici de culoare dintr-o imagine trebuie să înceapă cu reducerea semnificativă a numărului de culori existente în imagine. Acest lucru se realizează prin divizarea spațiului de culoare. Putem exprima formal acest lucru prin trecerea, pe baza unei relații de divizare Q , de la un spațiu de culoare C cu n culori posibile la un subspațiu de culoare P cu doar m culori posibile,

$$P = \{c_1, c_2, \dots, c_m \mid c_i \in C, m \ll n\}: Q: C \rightarrow P \quad (1)$$

Divizarea spațiului de culoare se poate realiza static sau dinamic. Dintre procedeele statice, cel mai simplu este luarea în considerare a unor biți cei mai semnificativi din octeții de codare a culorilor. O asemenea procedură nu are însă semnificație din punctul de vedere al percepției culorilor și produce modificări majore ale imaginilor din punct de vedere coloristic. Alte metode statice sunt: partiționarea spațiului de culoare, clusterizarea spațiului de culoare și utilizarea unor culori de referință. Procedeele dinamice implică utilizarea tehnicilor de segmentare a imaginilor [Schettini, 2001]¹².

Un exemplu clasic de reducere a numărului de culori este cel folosit de sistemul QBIC de la IBM, care face reducerea numărului de culori a celui mai uzual spațiu de culoare, RGB, codat pe 24 de biți, de la circa 1,68 milioane de culori la numai 4.096, prin divizarea fiecărui plan de culoare în numai 16 nivele, în locul celor 256 de nivele standard.

¹⁰ Idem.

¹¹ Stricker, M., Orengo, M., „Similarity of color images”, Proc. SPIE: Storage and Retrieval of Image and Video III, San Jose, CA, USA, 1995, pp. 381-392.

¹² Idem 8.

Există numeroase exemple de reducere a numărului de culori, la câteva sute de culori, la 64 de culori sau chiar la numai mult de 11 culori de bază. Cu cât se utilizează o metodă mai sofisticată, implicând, de exemplu, tehnici de segmentare, cu atât se obțin imagini mai apropiate de cea inițială din punct de vedere coloristic, deși numărul de culori utilizate scade drastic în toate cazurile. Ideal ar fi ca în urma acestor prelucrări să se obțină setul de culori dominante din imagine. Chiar și în cazul atingerii acestui ideal, aceste culori dominante ar fi foarte greu de utilizat ca și caracteristici de imagine, pentru că numărul lor ar varia de la o imagine la alta și, în plus, ele nu ar fi ordonate consistent pentru un set mare de imagini, ca să nu mai vorbim de cazul unui număr cvasi-infinit de imagini. Aceste două motive fac greu de comparat imaginile pe baza culorilor dominante, fiind foarte greu de găsit o măsură de tip distanță care să aprecieze gradul de similitudine a imaginilor pe baza lor.

2.2 Indecși de culoare bazați pe calculul histogramelor

Una dintre modalitățile cele mai uzuale de reprezentare a culorilor este calculul histogramei de culoare. Aceasta este un vector $H = [h_1, h_2, \dots, h_m]$ ale cărei componente h_k conțin numărul de pixeli de culoare k ce apar într-o imagine. Histograma de culoare este o caracteristică globală a unei imagini care estimează densitatea de probabilitate ca un pixel să aibă o anumită culoare.

Dacă avem două imagini I_1 și I_2 pentru care s-au calculat histogramele H^1 și H^2 bazate pe o aceeași paletă de m culori, gradul lor de similitudine poate fi apreciat, de exemplu, pe baza următoarei distanțe dintre histograme, numită intersecția histogramelor:

$$d(I_1, I_2) = d(H^1, H^2) = \frac{\sum_{k=1}^m \min(h_k^1, h_k^2)}{\sum_{k=1}^m h_k^2} \quad (2)$$

Dacă imaginile au și aceeași dimensiune (sau histogramele au fost scalate ca să aibă aceeași dimensiune) și sunt bazate pe aceeași paletă de culori, atunci se poate aplica direct suma pătratică a diferențelor (metrica L_2):

$$d(I_1, I_2) = d(H^1, H^2) = \sqrt{\sum_{k=1}^m (h_k^1 - h_k^2)^2} \quad (3)$$

sau chiar suma valorilor absolute a diferențelor (metrica L_1):

$$d(I_1, I_2) = d(H^1, H^2) = \sum_{k=1}^m |h_k^1 - h_k^2| \quad (4)$$

Rezultatele nu sunt însă prea bune la regăsirea imaginilor prin utilizarea acestor metrici, în principal din cauză că cele mai mici schimbări din histograme, datorate modificării condițiilor de iluminare, pot duce foarte ușor la mărirea distanțelor dintre imagini și la luarea unor decizii eronate de nesimilaritate. De

asemenea, histogramele de culoare ale unor imagini similare pot fi foarte îndepărtate din punctul de vedere al acestor metrice simple [Stricker, 1995]¹³. O oarecare îmbunătățire față de această situație se poate obține prin utilizarea histogramelor cumulative introduse de Stricker și Orego [Stricker, 1995]¹⁴, în care componentele histogramei cumulative $H_c(M)$ sunt date de relația bazată pe componentele h_i ale histogramei $H(M)$:

$$h_{ck} = \sum_{i=1}^{k-1} h_i \quad (5)$$

O altă distanță utilizată este distanța cosinus, de tipul:

$$d(I_1, I_2) = d(H^1, H^2) = \frac{2}{\pi} \cos^{-1} \left(\frac{\sum_{k=1}^m h_k^1 h_k^2}{\min(\|H^1\|, \|H^2\|)} \right) \quad (6)$$

Au fost imaginate și testate numeroase alte distanțe, bazate în principal pe codarea similitudinii perceptuale dintre culori (inclusiv prin metode fuzzy). O ilustrare simplă a acestei direcții este distanța propusă de Hafner [Hafner, 1995]¹⁵:

$$d(I_1, I_2) = d(H^1, H^2) = \sqrt{\sum_i \sum_j a_{ij} (h_i^1 - h_j^2)(h_i^1 - h_j^2)} \quad (7)$$

unde coeficienții a_{ij} codifică similitudinea dintre culorile i și j și se pot calcula pe baza unei expresii de tipul:

$$a_{ij} = 1 - \frac{d_{ij}}{\max_{ij}(d_{ij})} \quad (8)$$

cu d_{ij} fiind distanțe euclidiene între culori într-o variantă de spațiu de culoare de tip Munsell.

Există mai multe alte variante de definire a coeficienților a_{ij} , din ce în ce mai complicate și mai costisitoare din punctul de vedere al calculului lor, dar, la momentul utilizării lor în regăsirea imaginilor similare, toate se lovesc de problema neunicității histogramelor: aceeași histogramă poate corespunde la mai multe imagini (de exemplu modificarea poziției unui obiect într-o imagine nu schimbă histograma).

O variantă interesantă este și utilizarea unei distanțe Hausdorf modificate, care pentru două imagini I_1 și I_2 , fiecare caracterizate de setul de culori C^1 și C^2 , are expresia:

$$d(I_1, I_2) = H(C^1, C^2) = \max(\max_{c^1 \in C^1} \min_{c^2 \in C^2} \text{dist}(c^1, c^2), \max_{c^2 \in C^2} \min_{c^1 \in C^1} \text{dist}(c^1, c^2)) \quad (9)$$

din care rezultă distanța maximă dintre culorile din C^1 și C^2 (mai întâi se ia fiecare culoare din C^1 , se găsește culoarea din C^2 cea mai apropiată și se formează un set de

¹³ Stricker, M., Orego, M., „Similarity of color images”, Proc. SPIE: Storage and Retrieval of Image and Video III, San Jose, CA, USA, 1995, pp. 381-392.

¹⁴ Idem.

¹⁵ Hafner, J., Sawhney, H.S., Esquitz, W., Flickner, M., Niblack, W., „Efficient color histogram indexing for quadratic form distance functions”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 17, pp. 729-736, 1995

distanțe minime, din care se alege cea mai mare, după care se ia fiecare culoare din C^2 , se găsește culoarea din C^1 cea mai apropiată și se formează un al doilea set de distanțe minime, din care se alege din nou cea mai mare, la sfârșit selectându-se cea mai mare valoare din aceste două distanțe parțial alese).

O ultimă considerație referitoare la utilizarea histogramelor pentru calculul indecșilor de culoare: chiar dacă ar fi rezolvate problemele reliefate mai sus, problema alegerii optime a numărului de componente ale unei histogramme rămâne nerezolvată. Oricât de bună ar fi o anumită alegere, împărțirea plajei totale a histogrammei în elemente mai mici va produce efectele nedorite ale oricărei cuantificări.

2.3 Algoritmul centrului de greutate a culorilor (Chromatic barycenter algorithm)[Shiuh-sheng Yu, 1996]¹⁶

Centrul de greutate al culorilor pentru o regiune dintr-o imagine este media valorii componentelor de culoare ale pixelilor din acea regiune. De exemplu, în spațiul RGB, dacă avem o regiune perfect verde, centrul de greutate al culorilor pentru această regiune va fi (0,255,0).

Se poate defini o măsură a similarității s dintre două regiuni în funcție de distanța d dintre centrele lor de greutate a culorilor:

$$s(d) = \begin{cases} 100\%, & \text{dacă } d \leq T_1 \\ 100 \left(\frac{T_2 - d}{T_2 - T_1} \right) \% & \text{dacă } T_1 < d \leq T_2 \\ 0 \% & \text{dacă } d > T_2 \end{cases} \quad (10)$$

unde T_1 este stabilit în funcție de sensibilitatea ochiului uman, de exemplu 1/128 din mărimea spațiului de culoare (adică din numărul de nivele de gri luate în considerare pentru fiecare componentă de culoare), iar T_2 poate fi un parametru de precizie, de exemplu 1/8 din mărimea spațiului de culoare.

Este clar că nu putem compara două imagini doar pe baza unui singur centru de greutate al culorilor. De aceea imaginile sunt împărțite într-un număr de sub-regiuni, se calculează similaritățile dintre regiunile corespondente ale celor două imagini, de exemplu A și B , și se face media acestora:

$$Sim(A, B) = \frac{1}{n} \sum_{i=1}^n s(d_i(A, B)) \quad (11)$$

unde

$$d_i(A, B) = \|C_i^A - C_i^B\| \quad (12)$$

¹⁶ Shiuh-sheng Yu, Jinn-Rong Liou, Wen-Chin Chen, „Computational Similarity Based on Chromatic Barycenter Algorithm”, IEEE Transactions on Consumer Electronics, vol. 42, nr. 2, mai 1996.

iar C_i^A și C_i^B sunt centrele de greutate a culorilor ale regiunilor cu număr de ordine i din imaginile A și B .

În [Shiuh-sheng Yu, 1996]¹⁷ se arată că regiunile de 3x3 pixeli pot fi luate în considerare ca o variantă practică pentru acest algoritm. În plus, există variante care pot compensa distorsiunile de culoare, diferențele de rezoluție între imagini și deplasarea culorilor. S-au obținut valori foarte bune pentru rata de recunoaștere și pentru precizia recunoașterii. În plus, algoritmul se poate aplica și pe componenta continuă a imaginilor de tip JPEG, ceea ce înseamnă că acestea nu mai trebuie decodate complet. Aceasta deoarece componenta continuă a coeficienților DCT reprezintă chiar centrul de greutate al culorilor.

2.4 Momente de culoare

Ideea folosirii momentelor pentru caracterizarea distribuției culorilor într-o imagine digitală vine din teoria probabilității în care se știe că distribuția de probabilitate este caracterizată în mod unic de momentele sale, respectiv de momentele sale centrale [Stricker, 1995]¹⁸. Astfel, având în vedere că putem interpreta distribuția de culoare dintr-o imagine ca și distribuție de probabilitate, rezultă că putem caracteriza distribuția de culoare cu ajutorul momentelor. De exemplu, în cazul unei imagini RGB, separăm mai întâi planele de culoare R, G și B, și calculăm apoi în fiecare plan momentele de gradul 1, 2 și 3. Momentul de gradul 1 reprezintă media și deci astfel se pot determina nivelele medii de roșu, verde și albastru, care luate ca triplet ne vor da culoarea medie a imaginii digitale. Momentul de gradul 2 reprezintă varianța sau deviația standard, în timp ce momentul de gradul 3 reprezintă asimetria.

Pentru cazul general al unei imagini care poate fi separată în c planuri de culoare, iar pentru fiecare plan de culoare i avem valori ale pixelilor j ($j=1\dots N$) notate cu p_{ij} , atunci avem următoarele valori pentru momentele de gradul 1, 2 și 3 [Stricker, 1995]¹⁹:

$$M_i = \frac{1}{N} \sum_{j=1}^N p_{ij}, \quad V_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (p_{ij} - M_i)^2}, \quad A_i = \sqrt[3]{\frac{1}{N} \sum_{j=1}^N (p_{ij} - M_i)^3} \quad (13)$$

și se poate lua în considerare o măsură de similaritate de tipul:

$$d_{mom} = \sum_{i=1}^c \left(w_{i1} |M_i^A - M_i^B| + w_{i2} |V_i^A - V_i^B| + w_{i3} |A_i^A - A_i^B| \right) \quad (14)$$

în cazul a două imagini A și B , fiecare conținând N pixeli. Ponderile w_{ij} au fost introduse pentru a permite ajustarea similarității în funcție de specificul aplicațiilor la

¹⁷ Shiuh-sheng Yu, Jinn-Rong Liou, Wen-Chin Chen, „Computational Similarity Based on Chromatic Barycenter Algorithm”, IEEE Transactions on Consumer Electronics, vol. 42, nr. 2, mai 1996.

¹⁸ Stricker, M., Orengo, M., „Similarity of color images”, Proc. SPIE: Storage and Retrieval of Image and Video III, San Jose, CA, USA, 1995, pp. 381-392.

¹⁹ Stricker, M., Orengo, M., „Similarity of color images”, Proc. SPIE: Storage and Retrieval of Image and Video III, San Jose, CA, USA, 1995, pp. 381-392.

care este întrebuințată această metodă. Spre exemplu, dacă se știe apriori că toate imaginile au fost realizate în aceleași condiții de iluminare, putem penaliza diferențele de culoare medie prin considerarea unor ponderi w_{11} mai mari decât w_{12} și respectiv w_{13} .

Funcția d_{mom} este doar o măsură de similaritate deoarece s-au luat în considerare doar o mică parte dintre momentele distribuției de culoare și deci este foarte posibil ca două distribuții de culoare diferite să aibă aceleași momente de grad 1, 2 și 3 și să rezulte $d_{mom}=0$.

2.5 Imaginea proprie (*Eigen image* sau *Eigen vectors*)

Este o metodă foarte populară pentru discriminarea obiectelor de același tip și a fost utilizată mult la recunoașterea fețelor sub denumirea de *Eigen faces*. Practic, metoda face o proiecție a imaginii dintr-un spațiu cu mai multe dimensiuni într-un spațiu cu mai puține dimensiuni, ceea ce conduce la obținerea unor vectori proprii care realizează compactarea trăsăturilor unice ale imaginilor. Acești vectori proprii sunt apoi comparați unii cu alții în baza de date de imagini și se pot regăsi astfel imagini similare.

Înainte de a extrage vectorii proprii dintr-o imagine, este necesar ca aceasta să fie preprocesată, ceea ce include următoarele operațiuni: scăderea rezoluției imaginii de la 3 la 5 ori, decuparea precisă din imagine a obiectului de analizat (de exemplu un tablou dintr-o expoziție), identificarea unor puncte importante din imagine (cum ar fi colțurile unui tablou), transformarea de perspectivă a imaginii (cum ar fi cazul unui tablou fotografiat dintr-o parte și care trebuie adus în imagine ca și cum ar fi fotografiat perfect din față). Toate aceste preprocesări sunt necesare pentru că metoda este foarte sensibilă la diferențe de luminozitate și aspect existente între imaginile comparate.

O ultimă procesare constă în scăderea valorii medii a intensității pixelilor din imagine din valoarea fiecărui pixel. După această operație se generează o matrice S pe baza vectorilor de imagine f_i , câte unul pentru fiecare rând de pixeli din imagine :

$$S = \left(f_1 - \frac{1}{N} \sum f_1, \dots, f_m - \frac{1}{N} \sum f_m \right) \quad (15)$$

Imaginile proprii e_i sunt calculate apoi după metoda lui Sirovich și Kirby, mai întâi ca v_i :

$$S^H S = \lambda_i v_i, \forall i = 1 \dots M \quad (16)$$

și apoi ca:

$$e_i = S v_i \quad (17)$$

Aceste imagini proprii sunt calculate pentru componenta de luminanță a imaginii și pentru componenta de roșu, iar pentru compararea a două imagini s-a dovedit eficientă distanța euclidiană dintre acești vectori proprii, numiți *Eigen images*.

2.6 Vectori de coerență de culoare (Color coherence vectors) [Pass, 1996]²⁰

Coerența unei culori este definită ca și gradul în care pixelii din imagine corespunzători acelei culori fac parte dintr-o regiune relativ mare de pixeli de culoare similară. Regiunile de aceeași culoare dintr-o imagine sunt regiuni semnificative pentru caracterizarea și compararea imaginilor. Metoda coerenței de culoare vine să îmbunătățească metoda histogramelor de culoare, prin aceea că pixelii din culorile care apar în histograma sunt în plus împărțiți în două categorii, pixeli coerenți, care fac parte din regiuni omogene de culoare, și pixeli incoerenți, care sunt izolați printre alte culori.

Etapele de procesare parcurse de această metodă sunt următoarele:

- aplicarea unui filtru de tip medie pentru a elimina diferențele de culoare mici dintre pixelii adiacenți (cu luarea în considerare a conectivității de 8);
- discretizarea spațiului de culoare în n culori;
- clasificarea pixelilor din fiecare dintre cele n culori ca și coerenți sau incoerenți. Acest lucru se realizează pe baza calculării componentelor conectate, pentru fiecare culoare n găsiindu-se lanțurile de pixeli conectați de aceeași culoare. Se alege apoi un prag, astfel că un pixel dintr-o anumită culoare este coerent dacă face parte dintr-un lanț conectat de pixeli de aceeași culoare în număr mai mare decât acest prag. Toți pixelii care fac parte din lanțuri cu un număr de pixeli mai mic decât pragul prestabilit sunt considerați incoerenți.

În acest fel, dacă vectorul de trăsături în cazul utilizării histogramelor este (h_1, h_2, \dots, h_n) , în cazul metodei coerenței de culoare va fi $((a_1, b_1), (a_2, b_2), \dots, (a_n, b_n))$, unde a_i reprezintă numărul de pixeli coerenți din culoarea i , iar b_i reprezintă numărul de pixeli incoerenți din culoarea i . Evident că $a_i + b_i = h_i$.

Pentru compararea a două imagini I și I' se utilizează mărimea:

$$\Delta I = \sum_{i=1}^n (|a_i - a'_i| + |b_i - b'_i|) \quad (18)$$

diferită față de

$$\Delta H = \sum_{i=1}^n |(a_i + b_i) - (a'_i + b'_i)| \quad (19)$$

utilizată în cazul histogramelor. Se creează astfel o mai fină distincție între culorile corespondente aparținând la două imagini, în sensul că dacă avem același număr total de pixeli de aceeași culoare, dar într-o imagine ei sunt dispersați, pe când în cealaltă ei sunt grupați, diferența dintre imagini datorată acelei culori va fi mare, în timp ce în cazul histogramelor diferența datorată acelei culori ar fi fost nulă.

Metoda este în mod natural invariantă la orientare și poziție, în același timp nefiind foarte mare consumatoare de timp de calcul.

²⁰ Pass, Greg, Zabih, Ramin, Miller, Justin, „Comparing Images Using Color Coherence Vectors”, Computer Science Department, Cornell University, 1996.

3. Trăsături de textură

Noțiunea de textură se referă la proprietățile pe care le are suprafața sau structura unui obiect, așa cum se regăsește ea într-o imagine [Sonka, 2008]²¹. Deși este utilizată frecvent și este imediat acceptată intuitiv, nu are o definiție precisă, tocmai datorită variabilității sale. Se poate defini textura ca ceva care constă din elemente aflate într-o relație mutuală [Sonka, 2008]²². Textura este puternic dependentă de factorul de scală. Se constituie din elemente de textură sau primitive de textură, care poartă uneori denumirea de texeli (prin analogie cu pixelii, elementele primare ale unei imagini bidimensionale, sau voxelii, elementele primare ale unei imagini tridimensionale).

Analiza texturii are ca scop fie recunoașterea texturilor, fie recunoașterea formelor pe bază de texturi. De aceea, în mod normal, detectarea proprietăților de textură conduce la posibilitatea segmentării imaginilor în mai multe regiuni sau chiar a detectării unor forme în cadrul imaginilor digitale. Textura este descrisă uzual ca fină, rugoasă, granulată, netedă, noțiuni destul de greu de captat la nivel de recunoaștere automată pe calculator. Aceste caracteristici pot fi însă detectate dacă se analizează tonurile de gri și structura texturii. Tonurile de gri exploatează proprietățile de intensitate ale pixelilor, în timp ce structura texturii exprimă inter-relațiile spațiale dintre primitivele de textură.

În timp ce un pixel poate fi descris prin poziție și nivel de gri, o primitivă de textură este formată dintr-o grupare contiguă de pixeli cu anumite caracteristici de ton sau regionale și care poate fi descrisă prin nivelul de gri mediu, maxim sau minim, prin forma și dimensiunile sale etc. Relația spațială dintre primitive poate fi aleatorie, poate fi dependentă în perechi sau poate exista între un număr oarecare de primitive mutual dependente. Textura unei imagini poate fi apoi descrisă prin numărul și tipul de primitive, precum și prin relațiile spațiale existente între acestea [Sonka, 2008]²³.

Texturile pot fi clasificate în texturi fine, caracterizate prin primitive mici și diferențe mari de tonalitate între primitivele alăturate, sau texturi granulate, cu primitive mari, formate din mai mulți pixeli. De asemenea, există texturi mai puțin evidente, cu inter-relații mici între primitive, ce pot fi descrise adecvat cu ajutorul frecvențelor de apariție a anumitor tipuri de primitive într-o anumită zonă, și există texturi evidente, care pot fi descrise suficient de bine doar prin frecvența de apariție a perechilor de primitive inter-relate spațial într-un anumit mod.

Pentru descrierea texturilor există două abordări, pe cale statistică sau pe cale sintactică. Metodele statistice calculează anumite proprietăți și sunt potrivite pentru cazurile în care dimensiunile primitivelor de textură sunt comparabile cu dimensiunea pixelilor. Metodele sintactice și cele hibride (combinație între metodele statistice și cele sintactice) sunt mai adecvate în cazul în care texturile sunt formate

²¹ Sonka, Milan, Hlavac, Vaclav, Boyle, Roger, „*Image Processing, Analysis, and Machine Vision*”, International Student Edition, Thomson, 2008.

²² *Idem.*

²³ Sonka, Milan, Hlavac, Vaclav, Boyle, Roger, „*Image Processing, Analysis, and Machine Vision*”, International Student Edition, Thomson, 2008.

din primitive care pot fi etichetate, adică pot fi descrise prin mai multe caracteristici, nu numai pe baza tonurilor de gri, cum ar fi descrierea unei primitive prin forma ei.

Cercetările care au fost efectuate cu privire la capacitatea ochiului uman de a recunoaște texturile au arătat că la baza acestei capacități stau textonii, adică niște formațiuni alungite (dreptunghiuri, elipse, segmente de dreaptă, capete de segment, intersecții, colțuri) care pot fi detectate de vedere inconștient și foarte rapid, în timp ce relațiile poziționale dintre textonii alăturați sunt decelate mult mai târziu.

Din punctul de vedere al descrierii unei imagini prin vectori de trăsături în vederea determinării similitudinilor dintre imagini, este foarte clar că metodele statistice sunt mai adecvate, datorită descrierii rapide și globale a texturilor, având și o specificitate redusă. Metodele sintactice necesită operații laborioase (mari consumatoare de timp) pentru detectarea formelor și a relațiilor dintre primitivele de textură, iar specificitatea acestor operații este relativ mare, diferind mult de la un tip de imagine la altul. De aceea, ne vom rezuma în acest capitol doar în a prezenta cele mai importante caracteristici statistice de textură ce pot fi ușor calculate pentru o imagine digitală.

3.1 Metode statistice de descriere a texturii

3.1.1 Autocorelația [Sonka, 2008]²⁴

Dacă avem o imagine I de dimensiuni $M \times N$ și cu valori ale nivelelor de gri ale pixelilor notate cu p_{ij} , $i = 1 \dots M$, $j = 1 \dots N$, atunci se pot calcula o serie de coeficienți de autocorelație între pixelii aflați la distanțe Δx (pe direcția Ox) și Δy (pe direcția Oy) pe baza următoarei expresii:

$$C_{autocor}(\Delta x, \Delta y) = \frac{MN}{(M - \Delta x)(N - \Delta y)} \frac{\sum_{i=1}^{M-\Delta x} \sum_{j=1}^{N-\Delta y} p_{ij} p_{(i+\Delta x)(j+\Delta y)}}{\sum_{i=1}^M \sum_{j=1}^N p_{ij}^2} \quad (20)$$

Se pot alege mai multe valori pentru distanțele de autocorelație Δx și Δy , iar seria rezultată de coeficienți de autocorelație poate fi considerată ca și vector de trăsături de textură pentru imaginea I .

3.1.2 Matrice de apariții corelate [Sonka, 2008]²⁵

Această metodă se bazează pe apariția repetată a unor configurații de nivele de gri care caracterizează textura. În cazul texturilor fine, o astfel de configurație are o variație rapidă cu distanța, pe când în cazul texturilor mai grosiere, variația cu distanța este mai lentă. Din nou luăm în considerare imaginea I descrisă ca mai sus și

²⁴ *Idem.*

²⁵ Sonka, Milan, Hlavac, Vaclav, Boyle, Roger, „Image Processing, Analysis, and Machine Vision”, International Student Edition, Thomson, 2008.

considerăm o anumită zonă dreptunghiulară din ea (o fereastră de $m \times n$ pixeli). Apariția repetată a unei configurații de nivele de gri ce denotă o anumită textură poate fi descrisă printr-o matrice de frecvențe relative notată $P_{\phi,d}(g_1, g_2)$, care descrie cât de des apar în fereastră doi pixeli cu nivele de gri g_1 și g_2 , aflați la distanța d unul de celălalt, pe direcția ϕ .

Iată câteva exemple de astfel de matrice, pentru cazurile simple ale direcțiilor corespunzătoare unghiurilor de 0° (direcția orizontală), 45° , 90° (direcția verticală) și 135° , exprimate formal:

$$\begin{aligned} P_{0^\circ,d}(g_1, g_2) &= \text{cardinalul } \{[(k, l), (i, j)] \in (m \times n) \times (m \times n) : k - i = 0, |l - j| = d, p_{kl} = g_1, p_{ij} = g_2\} \\ P_{45^\circ,d}(g_1, g_2) &= \text{cardinalul } \{[(k, l), (i, j)] \in (m \times n) \times (m \times n) : (k - i = d, l - j = -d) \vee (k - i = -d, l - j = d), p_{kl} = g_1, p_{ij} = g_2\} \\ P_{90^\circ,d}(g_1, g_2) &= \text{cardinalul } \{[(k, l), (i, j)] \in (m \times n) \times (m \times n) : |k - i| = d, l - j = 0, p_{kl} = g_1, p_{ij} = g_2\} \\ P_{135^\circ,d}(g_1, g_2) &= \text{cardinalul } \{[(k, l), (i, j)] \in (m \times n) \times (m \times n) : (k - i = d, l - j = d) \vee (k - i = -d, l - j = -d), p_{kl} = g_1, p_{ij} = g_2\} \end{aligned} \quad (21)$$

Pentru extragerea unor trăsături de textură pot fi calculate anumite mărimi compuse, bazate pe matricele de apariții corelate rezultate pentru o serie de ferestre dintr-o imagine și pentru diverse direcții și distanțe între pixeli, cum ar fi:

1. **energia**, o măsură a omogenității unei imagini, calculată cu formula:

$$\sum_{g_1, g_2} P_{\phi,d}^2(g_1, g_2) \quad (22)$$

2. **entropia**, calculată cu formula:

$$\sum_{g_1, g_2} P_{\phi,d}(g_1, g_2) \log_2 P_{\phi,d}(g_1, g_2) \quad (23)$$

3. **probabilitatea maximă**, calculată cu formula:

$$\max_{g_1, g_2} P_{\phi,d}(g_1, g_2) \quad (24)$$

4. **contrastul**, o măsură a variațiilor locale în imagine, calculat uzual cu formula:

$$\begin{aligned} \sum_{g_1, g_2} |g_1 - g_2|^2 \log_2 P_{\phi,d}(g_1, g_2), \quad \text{ce poate fi generalizată ca} \\ \sum_{g_1, g_2} |g_1 - g_2|^k \log_2 P_{\phi,d}^k(g_1, g_2) \end{aligned} \quad (25)$$

5. **momentul diferenței inverse**, calculat uzual cu formula:

$$\sum_{g_1, g_2 : g_1 \neq g_2} \frac{P_{\phi,d}(g_1, g_2)}{|g_1 - g_2|^2}, \text{ generalizată ca } \sum_{g_1, g_2 : g_1 \neq g_2} \frac{P_{\phi,d}^k(g_1, g_2)}{|g_1 - g_2|^k} \quad (26)$$

6. **corelația**, o măsură a liniarității imaginii, dependentă de direcție, calculată cu formula:

$$\frac{\sum_{g_1, g_2} [(g_1 g_2) P_{\phi, d}(g_1, g_2)] - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (27)$$

unde, μ_x, μ_y sunt valorile medii:

$$\mu_x = \sum_{g_1} g_1 \sum_{g_2} P_{\phi, d}(g_1, g_2), \quad \mu_y = \sum_{g_2} g_2 \sum_{g_1} P_{\phi, d}(g_1, g_2) \quad (28)$$

iar σ_x, σ_y sunt deviațiile standard:

$$\sigma_x = \sum_{g_1} (g_1 - \mu_x)^2 \sum_{g_2} P_{\phi, d}(g_1, g_2), \quad \sigma_y = \sum_{g_2} (g_2 - \mu_y)^2 \sum_{g_1} P_{\phi, d}(g_1, g_2) \quad (29)$$

Practic, utilizarea acestor matrice de apariții corelate se bazează pe determinarea valorilor componente ale acestor matrice pentru mai multe direcții și mai multe distanțe, după care sunt calculate cele șase mărimi compuse prezentate mai sus. Se obține astfel o serie destul de mare de mărimi corelate care pot fi utilizate pentru compararea gradului de similaritate dintre imagini.

Metoda are însă dezavantaje mari în ceea ce privește volumul de calcul și cantitatea de memorie necesară pentru a duce la bun sfârșit toate calculele. Se poate recurge însă la reducerea nivelelor de gri dintr-o imagine până la 64 sau chiar 32, la adaptarea ferestrelor de aplicare în funcție de mărimea regiunilor omogene și, în plus, a fost dezvoltat un algoritm rapid de calcul al matricelor de apariții corelate.

3.1.3 Statistica muchiilor [Sonka, 2008]²⁶

Pentru descrierea texturilor dintr-o imagine poate fi utilizată și distribuția sau frecvența muchiilor din acea imagine. Pentru a detecta muchiile dintr-o imagine se pot folosi oricare dintre operatorii clasici, cum ar fi Roberts, Sobel, Prewitt, Canny etc. O variantă mai simplă este calcularea unei funcții de tip gradient, dependentă de distanță, pentru descrierea texturii, $g_{ij}(d)$:

$$g_{ij}(d) = |p_{ij} - p_{(i+d)j}| + |p_{ij} - p_{(i-d)j}| + |p_{ij} - p_{i(j+d)}| + |p_{ij} - p_{i(j-d)}| \quad (30)$$

Această funcție este similară cu negativul funcției de autocorelație, prezentând un minim acolo unde funcția de autocorelație este maximă și un maxim acolo unde autocorelația este minimă [Sonka, 2008]²⁷. Pentru aplicarea acestei metode se calculează funcția $g_{ij}(d)$ pentru toți pixelii dintr-o regiune texturată a unei imagini și pentru diferite distanțe d și apoi se rețin valorile medii ale gradientului obținute pentru fiecare distanță ca și caracteristică de textură pentru regiunea respectivă.

²⁶ Sonka, Milan, Hlavac, Vaclav, Boyle, Roger, „Image Processing, Analysis, and Machine Vision”, International Student Edition, Thomson, 2008.

²⁷ Idem.

Există o serie de proprietăți ale texturilor care pot fi, de asemenea, derivate din statistica de ordinul întâi (primele trei) sau doi (penultimele două) a distribuției muchiilor:

- a. **granularitatea** – dată de valoarea densității muchiilor detectate. Cu cât sunt mai multe muchii detectate într-o fereastră, cu atât textura este mai fină, și viceversa;
- b. **contrastul** – texturile cu contrastul mărit sunt caracterizate de dimensiuni mari ale muchiilor;
- c. **diversitatea** – gradul de diversitate a muchiilor măsurat ca entropie a histogramei dimensiunilor muchiilor;
- d. **directivitatea** – măsurată ca și entropie a histogramei direcțiilor muchiilor. Texturile care au anumite direcții mai pronunțate au un număr semnificativ și par de vârfuri, în timp ce texturile fără directivitate au această histogramă mai uniformă;
- e. **liniaritatea** – indicată de existența unor perechi de muchii cu aceeași direcție, poziționate una în continuarea celeilalte, la diferite distanțe;
- f. **periodicitatea** – indicată de existența unor perechi de muchii cu aceeași direcție sau de direcție contrară, poziționate în paralel una față de alta, la diferite distanțe;
- g. **dimensiunea**.

3.1.4 Trăsăturile Tamura [Tamura, 1978]²⁸

Tamura și Mori au propus în 1978 șase trăsături de textură care corespund percepției vizuale umane: granularitatea, contrastul, direcționalitatea, liniaritatea, regularitatea și rugozitatea. După testarea acestor șase trăsături, primele trei s-au dovedit a fi cele mai semnificative, în sensul că s-au corelat cel mai mult cu percepția umană.

Granularitatea se referă la dimensiunea elementelor de textură. Cu cât este mai mic coeficientul de granularitate, cu atât este mai fină textura. În cazul în care într-o imagine este prezentă o macrotextură cu dimensiuni mai mari, dar și o microtextură foarte fină, primează macrotextura ca trăsătură definitorie pentru acea imagine. De aceea, pentru calculul granularității unui pixel (i, j) de intensitate p_{ij} al unei imagini, se utilizează o serie de operatori cu dimensiune variabilă, aplicați pe vecinătăți ale pixelului de interes cu mărimi de $2^k \times 2^k$ pixeli, $k = 1, \dots, 5$ (adică 2×2 , 4×4 , 8×8 , 16×16 și 32×32 pixeli) astfel:

- a. se parcurge mai întâi toată imaginea și se calculează pentru fiecare din cele cinci vecinătăți ale oricărui pixel (i, j) o medie a intensității pixelilor

$$M_k(i, j) = \frac{1}{2^{2k}} \sum_{v=1}^{2^k} \sum_{w=1}^{2^k} p_{i-2^{k-1}+v, j-2^{k-1}+w} \quad (31)$$

- b. apoi se mai parcurge odată toată imaginea și se calculează pentru fiecare pixel (i, j) și pentru fiecare din cele cinci vecinătăți diferențele pe orizontală și pe

²⁸ Tamura, H., Mori, S., Yamawaki, T., „Textural Features Corresponding to Visual Perception”, IEEE Transaction on Systems, Man, and Cybernetics, nr. 8(6), 1978, pp. 460–472.

verticală între valorile medii calculate pentru vecinătățile adiacente vecinătății pixelului curent:

$$D_k^{oriz}(i, j) = |M_k(i + 2^{k-1}, j) - M_k(i - 2^{k-1}, j)| \text{ și} \quad (32)$$

$$D_k^{vert}(i, j) = |M_k(i, j + 2^{k-1}) - M_k(i, j - 2^{k-1})| \quad (33)$$

- c. se mai parcurge încă o dată imaginea și pentru fiecare pixel (i, j) se alege ca valoare pentru granularitate valoarea k a mărimii vecinătății pentru care s-a calculat cea mai mare diferență la punctul 2, luând în considerare ambele direcții:

$$G(i, j) = \arg \max_{k=1 \dots 5} \max_{d=oriz, vert} D_k^d(i, j) \quad (34)$$

Se poate face și o medie a granularității pentru toată imaginea, de genul

$$G_{img} = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N 2^{G(i, j)} \quad (35)$$

dar este mai util pentru vectorul de trăsături să avem o valoare a granularității pentru fiecare pixel din imagine.

Contrastul poate fi considerat o măsură a calității imaginii. El este influențat de domeniul dinamic de nivele de gri prezente într-o imagine, de polarizarea distribuției de alb și negru în cadrul histogramei de nivele de gri, de tăria muchiilor și de frecvența de repetiție a modelelor în cadrul unei imagini. Se calculează astfel:

$$C = \frac{\sigma}{\alpha_4^z} \quad (36)$$

unde

$$\alpha_4 = \frac{\mu_4}{\sigma^4} \quad (37)$$

în care μ_4 este momentul de ordinul patru

$$\mu_4 = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (p_{i, j} - \mu)^4 \quad (38)$$

pentru o imagine de $M \times N$ pixeli pentru care intensitatea medie a pixelilor este μ și deviația standard este σ , valoarea lui z fiind determinată experimental ca $1/4$. Pentru că ne interesează ca pentru fiecare pixel din imagine să avem o măsură a contrastului, se aplică formulele de mai sus pentru o vecinătate de tipul 15×15 în jurul fiecărui pixel.

Direcționalitatea nu este relevantă ca și direcție propriu-zisă, ci doar ca prezență în cadrul unei texturi (două texturi care diferă doar prin orientare sunt considerate a avea aceeași direcționalitate). Pentru a calcula direcționalitatea, se calculează mai întâi derivatele pe orizontală (Δ_{oriz}) și pe verticală (Δ_{vert}) pentru fiecare pixel din imagine. Se poate realiza acest lucru prin convoluția imaginii cu următorii operatori de dimensiunea 3×3 :

$$\begin{array}{ccc} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{array} \quad \text{și} \quad \begin{array}{ccc} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{array} \quad (39)$$

În loc de acești doi operatori se pot utiliza și operatorii Sobel.

Se calculează apoi direcționalitatea pentru fiecare pixel (i, j) din imagine:

$$\theta = \frac{\pi}{2} + \arctg \frac{\Delta_{\text{vert}}(i, j)}{\Delta_{\text{oriz}}(i, j)} \quad (40)$$

În final obținem un set de trei coeficienți, de granularitate, contrast și directivitate, pentru fiecare pixel din imagine și acest set mare de trăsături poate fi folosit fie direct ca și o imagine de tip RGB, fie sub forma unei histograme tridimensionale, pentru a determina similaritatea între mai multe imagini.

3.1.5 Măsurarea energiei texturilor [Sonka, 2008]²⁹

K.I. Laws a propus în 1979 caracterizarea texturilor prin determinarea nivelor medii de gri, a muchiilor, petelor și undelor pe care le prezintă o imagine texturată. Pentru asta a propus o metodă de calcul bazată pe trei vectori simpli:

- $L_3 = (1, 2, 1)$ – pentru calculul mediei;
- $E_3 = (-1, 0, 1)$ – pentru calculul diferențelor de ordinul întâi care relevă muchiile;
- $S_3 = (-1, 2, -1)$ – pentru calculul diferențelor de ordinul doi care relevă petele.

Se obțin apoi alți cinci vectori prin operația de convoluție a acestor vectori cu ei însuși și între ei:

- $L_5 = (1, 4, 6, 4, 1)$
- $E_5 = (-1, -2, 0, 2, 1)$
- $S_5 = (-1, 0, 2, 0, -1)$
- $R_5 = (1, -4, 6, -4, 1)$
- $W_5 = (-1, 2, 0, -2, -1)$

Prin multiplicarea acestor vectori unii cu alții se pot obține o serie de măști Laws de mărime 5×5 , de tipul:

$$L_5^T \times S_5 = \begin{bmatrix} -1 & 0 & 2 & 0 & -1 \\ -4 & 0 & 8 & 0 & -4 \\ -6 & 0 & 12 & 0 & -6 \\ -4 & 0 & 8 & 0 & -4 \\ -1 & 0 & 2 & 0 & -1 \end{bmatrix} \quad (41)$$

²⁹ Sonka, Milan, Hlavac, Vaclav, Boyle, Roger, „Image Processing, Analysis, and Machine Vision”, International Student Edition, Thomson, 2008.

Pentru obținerea de trăsături de textură se execută operația de convoluție a unor măști Laws de acest fel cu imaginile texturate și apoi se fac calcule statistice de energie pe imaginea rezultată. Rezultatele numerice ale statisticilor de energie sunt cuprinse în cadrul unor vectori de trăsături de textură pe baza cărora pot fi comparate imaginile și texturile între ele.

3.1.6 Dimensiunea fractală și lacunaritatea texturilor [Sonka, 2008]³⁰

Caracterizarea fractală a texturilor se bazează pe determinarea dimensiunii fractale și a lacunarității ca măsuri ale rugozității și granularității unei imagini cu nuanțe de gri. Din punct de vedere topologic, o imagine de acest tip are 3 dimensiuni, două dimensiuni fiind spațiale, iar a treia dimensiune este constituită din valoarea intensității la nivelul fiecărui pixel din imagine. Dimensiunea fractală D poate fi estimată din relația coeficientului H a lui Hurst:

$$H = T_D - D \quad (42)$$

unde T_D este dimensiunea topologică a unei imagini, care este egală cu 3.

Dacă notăm cu p_{ij} valoarea intensității imaginii în punctul (i, j) , atunci parametrul H poate fi estimat din relația:

$$E((\Delta p)^2) = c[(\Delta r)^H]^2 = c(\Delta r)^{6-2D} \quad (43)$$

unde $E()$ este un operator expectanță, $\Delta p = p_{ij} - p_{kl}$ este variația de intensitate de la pixelul (i, j) la pixelul (k, l) , iar $\Delta r = \|(i, j) - (k, l)\|$ este distanța dintre acești pixeli. Pentru a simplifica estimarea dimensiunii fractale este suficient să se utilizeze relația următoare:

$$E(|\Delta p|) = \kappa(\Delta r)^{3-D} \quad (44)$$

în care $\kappa = E(|\Delta p|)_{\Delta r=1}$. Revenind la notația cu $H = 3 - D$ și logaritmand se obține relația:

$$\log E(|\Delta p|) = \log \kappa + H \log(\Delta r) \quad (45)$$

Parametrul H poate fi obținut prin utilizarea regresiei liniare a celor mai mici pătrate pentru a estima panta curbei diferențelor de nivele de gri $gd(k)$ în funcție de k pe o scară dublu logaritmică. Pentru o imagine cu $M \times M$ pixeli a căror intensitate este p_{ij} , avem:

$$gd(k) = \frac{1}{\mu} \left(\sum_{i=0}^{M-1} \sum_{j=0}^{M-k-1} |p_{ij} - p_{i(j+k)}| + \sum_{i=0}^{M-k-1} \sum_{j=0}^{M-1} |p_{ij} - p_{(i+k),j}| \right), \quad (46)$$

în care $\mu = 2M(M - k - 1)$, iar scara k variază de la 1 la o valoare prestabilită s . După determinarea lui H , se calculează ușor și D , dimensiunea fractală. Mai trebuie

³⁰ Idem.

determinată și eroarea de aproximare a liniei de regresie pentru a dovedi că textura analizată este un fractal și, deci, poate fi descrisă efectiv cu această măsură fractală. Dacă dimensiunea fractală a unei texturi este mică, atunci acea textură este fină, iar dacă dimensiunea fractală rezultată este mare, atunci textura este mai grosieră [Sonka, 2008]³¹.

În cazul unor texturi de aceeași dimensiune fractală, dar care totuși apar ca diferite ochiului uman, se utilizează un alt parametru, lacunaritatea. Să presupunem că avem o mulțime fractală A și că $P(m)$ reprezintă probabilitatea ca să existe m puncte într-o fereastră pătrată de dimensiune L centrată în jurul oricărui punct din A . Dacă în fereastra pătrată pot fi maxim N puncte, se poate spune că:

$$\sum_{m=1}^N P(m) = 1 \quad (47)$$

Atunci, lacunaritatea λ este definită ca:

$$\lambda = \frac{M_2 - M^2}{M^2} \quad (48)$$

unde

$$M = \sum_{m=1}^N mP(m) \quad \text{și} \quad M_2 = \sum_{m=1}^N m^2 P(m) \quad (49)$$

Lacunaritatea este bazată pe calcule statistice de gradul 2 și are valori mici pentru texturile fine și valori mari pentru texturile grosiere [Sonka, 2008]³².

Pentru obținerea unui vector de trăsături pe baza parametrilor de dimensiune fractală și lacunaritate, se aplică o abordare multi-rezoluție. Se calculează acești doi parametri pentru imaginea considerată la rezoluția maximă, apoi pentru aceeași imagine, dar cu rezoluția redusă la jumătate, ș.a.m.d. până se atinge un anumit număr de trepte de rezoluție. Colecția de parametri calculați la diferite rezoluții se constituie într-un vector de trăsături ce poate fi utilizat pentru aprecierea gradului de similitudine între imagini.

3.1.7 Descrierea multiscalară a texturilor cu wavelets sau cu filtre Gabor

Este posibilă utilizarea wavelet-urilor și a filtrelor Gabor pentru descrierea texturilor din cauza faptului că texturile sunt dependente de scala la care sunt percepute. În [Sonka, 2008]³³ se face o descriere a aplicării unor cadre discrete de wavelet-uri pentru a extrage trăsături de textură dintr-o imagine. Trăsăturile utilizate sunt niște semnături ale energiei wavelet-urilor, sau derivate de ordinal 2 ale acestora. Din păcate, aceste trăsături nu sunt invariante la translație, ceea ce le face

³¹ Sonka, Milan, Hlavac, Vaclav, Boyle, Roger, „Image Processing, Analysis, and Machine Vision”, International Student Edition, Thomson, 2008.

³² Idem.

³³ Sonka, Milan, Hlavac, Vaclav, Boyle, Roger, „Image Processing, Analysis, and Machine Vision”, International Student Edition, Thomson, 2008.

foarte dificil de utilizat. Ar fi necesar ca să se extragă trăsăturile din mai multe variante translatate ale imaginii originale, pentru a se depăși această problemă. Un alt inconvenient este însă volumul mare de calcul necesar în comparație cu alte metode, fiind aproape exclusă utilizarea în cazul colecțiilor foarte mari de imagini.

3.2 Metode sintactice de descriere a texturii [Sonka, 2008]³⁴

Descrierea sintactică a texturilor se bazează pe analogia existentă între relațiile spațiale ale primitivelor de textură și structura unui limbaj formal. Texturile sunt compuse din primitive care sunt dispuse aproape regulat unele față de celelalte. Una dintre cele mai eficiente metode de a descrie structura texturilor se constituie prin găsirea unei gramatici care conține regulile de compunere a texturilor din primitive, prin aplicarea unor reguli de transformare asupra unui set limitat de simboluri. Aceste simboluri sunt de fapt tipuri de primitive de textură și regulile de transformare reprezintă de fapt relațiile spațiale dintre aceste primitive. În realitate însă, texturile nu sunt chiar atât de regulate, ceea ce implică generarea unor reguli care să nu fie stricte, ci să aibă un oarecare grad de variabilitate.

Exemple de gramatici utilizate pentru descrierea texturilor sunt gramaticile de înlănțuire a formelor și gramaticile de graf, existând și gramatici de tip arbore sau de tip matrice.

O altă abordare constă în gruparea primitivelor în texturi ierarhice.

În fine, pot fi amintite aici și încercările de a utiliza metode combinate, hibride, în care se păstrează definiția precisă a primitivelor de textură, dar se recurge la metode statistice atunci când se iau în considerare relațiile de spațialitate dintre aceste primitive.

Aceste metode sunt destul de laborioase, necesită o analiză amănunțită și de înalt nivel al texturilor și nu oferă posibilitatea de a produce un vector de trăsături care să caracterizeze texturile în vederea comparării imaginilor în care ele apar.

4. Trăsături de formă

Formele sunt trăsături de nivel înalt care conțin informație semantică, spre deosebire de caracteristicile de bază de culoare și textură, care nu poartă o astfel de informație decât în cazuri particulare. Formele nu sunt caracteristici generale ale imaginilor, ci mai degrabă caracteristici ale obiectelor ce pot fi regăsite în imagini. Iar dificultatea constă în extragerea formelor obiectelor dintr-o imagine, proces numit segmentare, pentru că este foarte greu ca acest lucru să fie făcut cu precizie doar pe baza caracteristicilor de bază. Chiar și segmentarea imaginilor în regiuni de interes este foarte dificilă.

Dificultatea segmentării obiectelor provine și de la faptul că obiectele sunt în general tridimensionale în lumea reală, fiind redată doar bidimensional în imagini. Mai mult decât atât, compararea formelor unele cu altele este o problemă foarte dificilă.

³⁴ *Idem.*

Pentru descrierea formelor se merge pe două căi: descrierea conturului (deci numai a marginilor obiectului respectiv) și descrierea regiunii (descrierea obiectului ca un tot). Conturul unei forme poate fi descris pe baza încadrării lui într-un poligon sau într-o linie curbă închisă care să conțină în întregime obiectul respectiv și care la rândul lor pot fi descrise pe baza unei înlanțuiri de coordonate de pixeli, segmente de dreaptă, arce de cerc sau spline. Regiunile pot fi descrise pe baza unor seturi de puncte, a unor dreptunghiuri de încadrare sau a altor tipuri de poligoane de încadrare, discuri, șabloane deformabile sau scheletizări. Un caz interesant de descriere a formelor pe bază de regiuni este cel prezentat în [Carson, 2002]³⁵, Blobworld, unde regiunile de interes sunt reprezentate de forme eliptice determinate de centroidul obiectelor și de matricea lor de împrăștiere, la care se adaugă proprietățile de textură și două culori dominante din respectivele regiuni.

Se mai pot adăuga și caracteristicile legate de relaționarea spațială a obiectelor într-o imagine. Există relații de tip topologic, de exemplu, un obiect este „aproape” de altul, este „inclus” în altul sau este „adiacent” altui obiect, precum și relații determinate de poziția relativă a obiectelor, cum ar fi un obiect „în fața” altuia, în „dreapta” altuia sau „peste” alt obiect. Aceste tipuri de relații spațiale sunt cel mai bine descrise pe baza unor grafuri cu attribute, în care nodurile sunt obiectele, iar arcele care unesc nodurile între ele reprezintă o anumită relație între obiectele corespunzătoare nodurilor.

4.1. Descriptorii de formă din standardul MPEG 7

În cadrul standardului MPEG 7 au fost selectați o serie de descriptori de formă, criteriile de selecție a acestora fiind gradul lor de compactitate, invarianța la scalare, rotație și translație, precum și invarianța la distorsiunea formei datorată condițiilor de filmare.

Primul descriptor selectat se referă la suprafețe tridimensionale și nu interesează foarte mult acest studiu.

Al doilea descriptor este bazat pe regiuni și se rezumă la calcularea unor momente ce descriu distribuția spațială a regiunilor ce definesc obiecte. Se utilizează transformarea radială unghiulară (ART):

$$G_{nm} = \frac{c_n}{\pi} \int_0^{2\pi} \int_0^1 p_{\rho\theta} \cos(\pi n \rho) e^{-jm\theta} d\rho d\theta \quad (50)$$

unde $p_{\rho\theta}$ este valoarea intensității imaginii la distanța ρ de centrul cercului unitar și la unghiul θ , $c_0 = 1$ și $c_n = 2$, $n > 0$. Se rețin primele 35 de momente polare (până la ordinul 10), codificate pe 4 biți, pentru a se ajunge la un vector de trăsături de 140 de biți.

Un al treilea descriptor este bazat pe contur, mai precis pe spațiul scalar al curbării (CSS). În acest caz vectorul de trăsături este în medie de 112 biți, dintre care 6 biți pentru escentricitatea și circularitatea contururilor inițiale și finale, 6 biți pentru

³⁵ Carson, Chad, Belongie, Serge, Greenspan, Hayit, and Malik, Jitendra, „Blobworld: Image Segmentation using Expectation-Maximization and its Application to Image Querying”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, nr. 8, august 2002.

numărul de vârfuri din imaginea CSS, 7 biți pentru înălțimea celui mai mare vârf și apoi câte 9 biți pentru poziția celorlalte vârfuri.

Cel de-al patrulea descriptor combină descriptori bidimensionali ai imaginilor aceluiasi obiect tridimensional luate din diferite unghiuri.

4.2 Cazul Blobworld [Carson, 2002]³⁶

Un exemplu foarte interesant de utilizare a formelor pentru regăsirea imaginilor similare este sistemul Blobworld, bazat pe segmentarea imaginilor în regiuni care de multe ori corespund obiectelor din imagine. Reprezentarea Blobworld a unei imagini este obținută în urma parcurgerii a trei etape:

1. selecția unei scări potrivite pentru fiecare pixel din imagine și extragerea caracteristicilor de culoare, textură și poziție pentru acel pixel și la acea scară;
2. gruparea pixelilor în regiuni prin modelarea distribuției trăsăturilor pixelilor cu o mixtură de curbe Gauss și utilizarea maximizării expectanței;
3. fiecărei regiuni rezultate i se atașează o descriere globală a culorii și texturii acesteia, pentru a putea fi folosită în interogări.

În ceea ce privește caracteristicile de culoare, sistemul Blobworld folosește spațiul de culoare $L^*a^*b^*$, datorită faptului că acesta este aproape uniform din punct de vedere perceptual, rezultând că distanțele euclidiene dintre coeficienții (a,b) din același plan ab au semnificație de distanțe între culorile corespunzătoare. Cu toate acestea, sistemul nu reține caracteristici de culoare sub forma unor vectori de trăsături, ci realizează doar o netezire a fiecărui plan L , a , și b pentru a obține o reducere a numărului de culori, ceea ce favorizează segmentarea imaginii în mai puține regiuni relativ compacte. Se reține pentru fiecare pixel din imagine tripleta (L,a,b) care îi corespunde după netezire.

Pentru determinarea caracteristicilor de textură a fiecărui pixel din imagine se pune accentul pe stabilirea scalei la care se judecă această textură, având în vedere cât de important este acest factor pentru acest tip de caracteristică. Procedura de determinare a scalei este destul de complicată, bazată pe mărimea unei ferestre Gausiene în care se calculează și pentru care rezultă un gradient uniform al intensității pixelilor. Determinarea scalei de textură a unui pixel se realizează pe baza proprietății locale a unei imagini numită polaritate, care este o măsură a gradului în care vectorii de gradient al intensității imaginii sunt îndreptați în același direcție într-o anumită vecinătate. Prin inspectarea polarității, un pixel poate fi clasificat ca aparținând unei regiuni de tip muchie, de tip textură 1D sau 2D, sau aparținând unei regiuni uniforme, adică fără o textură anume. Găsindu-se scala pentru care un pixel poate fi parte a uneia dintre cele trei tipuri de regiuni specificate astfel, se rețin apoi pentru fiecare pixel caracteristicile de textură denumite polaritate, anizotropie și contrast normalizat.

La finalul acestor procesări, pentru fiecare pixel dintr-o imagine se obțin trei parametri de culoare (L,a,b) după netezire, trei parametri de textură și se mai includ în vectorul de trăsături al fiecărui pixel și poziția (x,y) a acestuia. Rezultă un set

³⁶ *Idem.*

foarte mare de parametri (8-dimensional) pentru o imagine, care este restrâns folosind algoritmul de maximizare a expectanței (Expectation-Maximization) pentru determinarea parametrilor, care maximizează asemănarea spațiului de caracteristici cu o mixtură de Gaussiene K . Acest algoritm ajută sistemul Blobworld să clasifice pixelii unei imagini prin clusterizare în regiuni cât mai reprezentative pentru acea imagine.

După ce fiecare pixel a fost inclus într-un anumit cluster, se trece la gruparea spațială a pixelilor care fac parte din fiecare cluster. Se face mai întâi etichetarea pixelilor conform clusterelor de care aparțin, după care se aplică un algoritm de conectare a componentelor cu aceeași etichetă. Se mai aplică și un pas de postprocesare pentru a demarca și alinia cât mai bine marginile regiunilor rezultate. Întreaga operațiune de segmentare pentru o imagine durează însă relativ mult, de ordinul minutelor pentru o imagine și pe un PC de la nivelul anului 2002, mult mai puțin acum, datorită progreselor obținute pe linia vitezei de calcul. Evident că autorii au testat procesarea paralelă a imaginilor tot la nivelul anului 2002.

Rezultatele practice prezentate de autori sunt interesante, reliefând faptul că acest tip de segmentare a imaginilor duce la determinarea unor regiuni din imagine reprezentând obiecte importante pentru asemănarea sau discriminarea lor față de alte imagini. Se obțin și unele rezultate nedorite, cum ar fi împărțirea fundalului în mai multe regiuni, faptul că marginile regiunilor nu corespund perfect marginilor obiectelor din imagine sau chiar faptul că obiectele din imagine sunt ratate complet, sunt împărțite între regiuni diferite sau sunt alipite cu regiuni mai mari.

În ceea ce privește sistemul Blobworld de regăsire a imaginilor, o imagine de test nu este transmisă pur și simplu sistemului spre căutare de imagini similare, ci mai întâi este supusă algoritmului de segmentare și extragere de trăsături, rezultând o imagine segmentată care este prezentată utilizatorului. Acesta poate apoi să selecteze din imagine regiunile care îl interesează dintre cele segmentate automat și poate de asemenea să stabilească importanța relativă a regiunilor alese, dacă sunt mai multe. Abia apoi se transmite din nou sistemului Blobworld comanda de căutare a imaginilor asemănătoare și se generează o listă de imagini asemănătoare pe baza distanțelor dintre regiunile de imagine selectate și regiunile imaginilor existente în baza de date.

Și rezultatele privitoare la imaginile regăsite atestă validitatea sistemului Blobworld și îl fac una din soluțiile interesante bazate pe analiza comparativă a regiunilor în care este segmentată o imagine, și nu direct pe baza unui vector de trăsături de culoare sau textură. Prețul acestui sistem este însă complexitatea calculelor necesare atât pentru segmentarea imaginii de test, cât și pentru segmentarea tuturor imaginilor din baza de date a unui astfel de sistem.

5. Brevet de invenție S.U.A. [Yagnik, 2011]³⁷

Interesant pentru acest studiu este brevetul de invenție nr. 8.065.313, „Metodă și aparatură pentru adnotarea automată a imaginilor” depus în Statele Unite ale Americii la 24.07.2006 de către Jay N. Yagnik și atribuit la 22.11.2011 companiei Google. Sistemul care face obiectul brevetului de invenție realizează citirea unei imagini, extragerea de trăsături caracteristice acestei imagini și identificarea unor alte imagini de pe Internet cu trăsături caracteristice similare. În plus, sistemul găsește informațiile de tip text (cuvintele cheie) asociate imaginilor similare celei de intrare și face o intersecție a cuvintelor cheie pentru a produce textul care este asociat automat imaginii de intrare. Sunt nu mai puțin de 38 de submetode pentru care se pretinde dreptul de inventator și le vom comenta doar pe acelea care sunt strict pe tematica acestui capitol.

În primul rând, înainte de a extrage trăsăturile din imaginea de intrare, care poate fi o imagine propriu-zisă codată JPEG, GIF, BMP, TIFF etc. sau poate fi la rândul ei extrasă dintr-o secvență video codată MPEG, sistemul împarte imaginea în mai multe blocuri dreptunghiulare. Trăsăturile sunt extrase pentru fiecare bloc în parte și este prevăzută și situația în care există o serie de blocuri de mai multe dimensiuni, iar imaginea de intrare este împărțită succesiv în blocuri de aceeași dimensiune până când se termină toate dimensiunile de blocuri din listă. În acest caz sunt extrase trăsături pentru fiecare bloc și fiecare dimensiune.

Autorul se referă apoi la tipul de trăsături care sunt extrase din fiecare bloc de imagine, și anume trăsături de formă, culoare și textură. Pentru trăsăturile de culoare se dau ca exemplu: histogramele de culoare, dispunerea histogramei de culoare și momentele de culoare. Între alte tehnici de procesare a imaginilor din care rezultă trăsături, autorul amintește transformata cosinus directă, analiza componentelor principale și wavelet-urile Gabor. De asemenea, este inclusă și metoda histogramei de orientare a muchiilor.

Fiecare dintre aceste trăsături sunt extrase pentru fiecare bloc de imagine, la diferite scări ale acestuia, și se creează un vector de trăsături complex cu toate aceste trăsături. Folosirea blocurilor de imagine de dimensiune variabilă este necesară pentru a putea caracteriza imaginea de intrare la diferite scări de percepere a acesteia, favorizând atât evidențierea detaliilor, dar detectând și informații globale despre imagine.

Pentru a găsi imagini similare cu imaginea de intrare, sistemul dispune de o multitudine de imagini aflate fie în baze de imagini proprii, fie culese de pe Internet, căroră li se aplică același algoritm de extragere a trăsăturilor, numai acestea fiind salvate într-o bază de date de trăsături. Această bază de date de trăsături este apoi utilizată pentru a găsi imaginile similare pe principiul găsirii acelor imagini care au trăsăturile cât mai apropiate de imaginea de intrare. Autorul amintește că o metodă simplă este aceea de contorizare a numărului de trăsături similare sau identice, dar sugerează și folosirea unor tehnici mai complexe referitoare la distribuțiile cu mai multe variabile, cum ar fi distribuția Gaussiană cu mai multe variabile.

³⁷ Yagnik, Jay N., „Method and Apparatus for Automatically Annotating Images”, brevet de invenție S.U.A. nr 8.065.313 din 22 noiembrie 2011.

După ce au fost identificate o serie de imagini similare, sistemul preia porțiunile de text care sunt alăturate imaginii, de exemplu în cazul imaginilor preluate din pagini de Internet. Autorul se bazează pe presupunerea că textul alăturat unei imagini de pe Internet este cel mai probabil să fie relevant pentru imaginea respectivă. Din aceste texte sunt extrase cuvinte cheie, care pot sau nu să fie relevante pentru imaginea respectivă.

Colecția de cuvinte cheie astfel obținută este îmbogățită suplimentar prin adăugarea unor sinonime pe baza unor dicționare tezaur încorporate în sistem. Pentru fiecare imagine similară cu cea de la intrare se găsește astfel câte un set de cuvinte cheie. Ultimul pas constă în efectuarea unei intersecții a acestor seturi de cuvinte cheie, pentru a le găsi pe cele care apar în toate seturile. Cuvintele cheie rezultate după intersecție sunt atribuite imaginii de la intrare, fiind considerate ca o descriere a acesteia.

Concluzii

S-au inventariat metode de generare a unor vectori de trăsături din imaginile digitale, capabili să asigure regăsirea unor imagini asemănătoare în colecții masive de imagini. Întrucât metodele de extragere a trăsăturilor bazate pe caracteristicile de culoare ale imaginilor digitale sunt cele mai directe, cele mai simple și cele mai puțin costisitoare din punct de vedere al timpului de calcul, cea mai mare parte a capitolului a fost consacrată acestora. Ele au fost prezentate în ordine, de la cele mai simple la cele mai complicate, împreună cu formulele de calcul aferente și cu distanțele sau măsurile de similaritate adecvate fiecărei metode. Deși prezintă numeroase avantaje de aplicativitate, vectorii de trăsături rezultați în urma analizei culorilor dintr-o imagine prezintă o limitare majoră, dată de imposibilitatea captării automate a informației semantice din imagine doar pe baza distribuției culorilor. Acest decalaj semantic face ca să fie aproape imposibil de estimat chiar și până la ce procente maxime de regăsire a imaginilor similare se poate ajunge pe baza caracteristicilor de culoare. Varianta controlului uman în acest caz nu este de folos, pentru că, pe de o parte omul nu poate estima foarte precis distribuția de culoare dintr-o imagine, iar pe de altă parte, în momentul în care unui om i se prezintă o imagine, informația semantică este inconștient receptată.

Celelalte două tipuri de trăsături, de textură și de formă, au fost tratate în mai puțin detaliu. Deși metodele statistice de caracterizare a texturilor generează vectori de trăsături similari cu cei rezultați prin analiza culorilor, texturile nu sunt atât de utile în măsurarea asemănării dintre imagini. Texturile se referă la anumite regiuni din imagini, nu pot fi descrise foarte precis și, în plus, necesită și un timp de calcul mult mai mare. Mai degrabă texturile ar putea fi utilizate într-un pas de rafinare a numărului de imagini similare regăsite pe baza trăsăturilor de culoare.

În fine, trăsăturile de formă sunt cele mai greu de determinat și de încapsulat în vectori de trăsături potriviți pentru sistemele de regăsire a imaginilor. Formele poartă informație semantică și detectarea lor poate duce la descoperirea obiectelor dintr-o imagine și, într-un pas ulterior, la interpretarea imaginilor. Principala metodă

prin care se poate ajunge la detectarea formelor este segmentarea, care necesită însă timp și condiții de aplicare care ies din sfera sistemelor de regăsire automată a imaginilor similare. Cu toate acestea, am trecut în revistă un sistem bazat pe caracteristicile de formă și cu care s-au obținut rezultate interesante la nivelul anului 2002.

Studiul se încheie cu prezentarea sumară a unui brevet de invenție atribuit companiei Google în 2011 și care se încadrează foarte bine în domeniul aflat în discuție.

Bibliografie

- Datta, Ritendra, Joshi, Dhiraj, Li, Jia and Wang, James Z., „Image Retrieval: Ideas, Influences, and Trends of the New Age”, The Pennsylvania State University, ACM Computing Surveys, Vol. 40, No. 2, Article 5, Publication date: April 2008.
- Carson, Chad, Belongie, Serge, Greenspan, Hayit, and Malik, Jitendra, „Blobworld: Image Segmentation using Expectation-Maximization and its Application to Image Querying”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, nr. 8, august 2002
- Eakins, John, Graham, Margaret, „JISC Technology Applications Programme Report #39”, University of Northumbria at Newcastle, October 1999.
- Gonzalez, Rafael C., Woods, Richard E., „Digital Image Processing”, Prentice Hall, Upper Saddle River, NJ, 2002.
- Gonzalez, Rafael C., Woods, Richard E., and Eddins, Steven L., „Digital Image Processing using MATLAB”, Gatesmark Publishing, 2009.
- Hafner, J., Sawhney, H.S., Esquitz, W., Flickner, M., Niblack, W., „Efficient color histogram indexing for quadratic form distance functions”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 17, pp. 729-736, 1995
- Pass, Greg, Zabih, Ramin, Miller, Justin, „Comparing Images Using Color Coherence Vectors”, Computer Science Department, Cornell University, 1996
- Pratt, William K. „Digital Image Processing”, John Wiley & Sons, 2001.
- Schettini, Raimondo, Ciocca, Gianluigi, Zuffi, Silvia, „Indexing and Retrieval in Colour Image Databases”, Chapter 10 in Colour Image Science – Exploiting Digital Media, Editori Lindsay W. MacDonald, M. Ronnier Luo, John Wiley & Sons, England, 2001, pp. 183 – 213
- Shiuh-sheng Yu, Jinn-Rong Liou, Wen-Chin Chen, „Computational Similarity Based on Chromatic Barycenter Algorithm”, IEEE Transactions on Consumer Electronics, Vol. 42, nr. 2, mai 1996
- Sonka, Milan, Hlavac, Vaclav, Boyle, Roger, „Image Processing, Analysis, and Machine Vision”, International Student Edition, Thomson, 2008.
- Stricker, M., Orengo, M., „Similarity of color images”, Proc. SPIE: Storage and Retrieval of Image and Video III, San Jose, CA, USA, 1995, pp. 381-392
- Tamura, H., Mori, S., Yamawaki, T., „Textural Features Corresponding to Visual Perception”, IEEE Transaction on Systems, Man, and Cybernetics, nr. 8(6), 1978, pp. 460-472
- Veltkamp, Remco C., Tanase, Mirela, „Content-Based Image Retrieval Systems: A Survey, Department of Computing Science”, Utrecht University, Technical Report UU-CS-2000-34, October 28, 2002.
- Yagnik, Jay N., „Method and Apparatus for Automatically Annotating Images”, brevet de invenție S.U.A. nr 8.065.313 din 22 noiembrie 2011

De la percepția vizuală către detecția automată a texturii

MIHAELA COSTIN¹, ANCA IGNAT²

ABSTRACT. Modeling characteristics close to human and primate vision in order to conceive functions and programs for artificial vision is a challenging research. Neurosciences are revealing important aspects on vision from the cognitive point of view. We are trying to formalise the modularity of the visual system in the attempt of automatic recognition of the surrounding environment. Hierarchic³ processing of visual features in a similar manner as humans is a natural way of treating information upon its importance in the perception of the living beings. Forms, colors and textures are among the most important features that are perceived at first glance. Dual Tree Complex Wavelet Transform, DTCWT prove to be very reliable and suitable in detecting texture features. Similarities among textures might be measured using appropriate distances. An important number of trial and error essays on texture detection, by this new method, have been conducted. We are describing them, reporting and discussing the results.

KEYWORDS: *texture, vision, Dual Tree Complex Wavelet Transform, DTCWT.*

1. Introducere – percepție vizuală - trăsături caracteristice

Cu scopul de a formaliza și modela abilitățile vizuale umane și de a le transpune în programe care să coordoneze vederea artificială în special pentru domeniul roboticii, este important să studiem din punct de vedere cognitiv metodele prin care putem furniza cea mai mare cantitate de informații sistemului. Sistemul vizual uman include retina, cu receptorii sensibili pentru lumină, căile de transmitere a semnalelor nervoase și zonele de proiecție corticale.

Studiile neurocognitive⁴ au demonstrat similaritatea unor module din sistemul vizual uman cu cel al primatelor și al altor mamifere. Premiul Nobel în fiziologie și

¹ Mihaela Costin, Academia Romana filiala Iași, studiu post-doctoral în cadrul proiectului „Societatea Bazată pe Cunoaștere – cercetări, dezbateri, perspective”, cofinanțat de Uniunea Europeană și Guvernul României din Fondul Social European prin Programul Operațional Sectorial Dezvoltarea Resurselor Umane 2007–2013, POSDRU 56815.

² Anca Ignat, Facultatea de Informatică, Universitatea „Al. I. Cuza”, Iași.

³ Niko Wilbert, *Hierarchical Slow Feature Analysis on Visual Stimuli and Top-Down Reconstruction*, Doctoral thesis, Humboldt Univ., Berlin, <http://www.ini.rub.de/PEOPLE/wiskott/Reprints/Wilbert-2012-PhDThesis.pdf>.

⁴ Michele Fabre-Thorpe, Arnaud Delorme et Ghislaine Richard, *Singes et Hommes face au monde visuel: la Catégorisation*, *Primatologie*, 2, 111-139, 1999.

medicină din anul 1981 a fost acordat cercetătorilor David Hubel și Torsten Wiesel pentru contribuțiile aduse în cunoașterea sistemului vizual. În 1958 ei realizaseră experimente care demonstau existența în cortexul vizual al felinei, a unor coloane (grupuri) de celule sensibile la orientarea unghiulară a liniilor prezente în câmpul vizual, (grupuri de celule nervoase care prezintă selectivitate la orientare). După cum se va vedea pe parcursul acestui capitol, orientarea unghiulară are o importanță deosebită în identificarea particularităților (trăsăturilor caracteristice) ale imaginilor și în particular, a texturii.

Textura ocupă un loc important pentru extragerea de indicii necesare în formalizarea vederii artificiale. Obiectivele majore în studiul texturii sunt înțelegerea, modelarea și procesarea texturii pentru a simula procesul uman de învățare vizuală, cu ajutorul tehnologiilor informatice.

Experimentele realizate cu ajutorul primatelor (Simon Thorpe)⁵ au demonstrat că vederea se focalizează mai întâi pe figurile umane, pe ființele care apar în cadru, într-un timp înregistrat de 20 ms, până la 150 ms. Doar într-o a doua etapă se remarcă categoria secundară din imagine, ținând cont de background, de fundalul imaginii, care adeseori prezintă texturi ce constituie indicii importante pentru înțelegerea scenei, a contextului.

Studii referitoare la existența reprezentărilor vizuale comune atât pentru om cât și pentru primat⁶, unele dintre ele efectuate în cadrul laboratorului de cercetări CERCO – Centre de Recherche Cerveau & Cognition – Université Toulouse III, conduc la unele concluzii din care se pot deduce elemente despre percepția zonelor de textură. Utilizând imagini ale unor obiecte manufacturate de om, sau imagini de animale în contextul unor scene specific umane (create de om) sau în context natural, studiul relevă faptul că inițial macacii testați și-au bazat deciziile categoriale pe prezența animalelor în scenă, fără a ține cont de background. Totuși, ca și la om, primatele au dat răspunsuri mai bune în clasificarea unor asociații de categorii congruente obiect/context, în special atunci când dimensiunea mică a obiectelor a favorizat informația de background. Îmbunătățirea acurateței și îmbunătățirea vitezei de răspuns a putut fi atribuită congruenței categoriilor superordonate (de genul obiect vizualizat din categoria hrană – acțiune coordonată – a se hrăni, a mânca) cu rezultate ale testelor uimitor de similare atât pentru primat cât și pentru subiecții umani în experimente care au utilizat aceleași teste și aceiași stimuli⁷.

Aceste rezultate sugerează o procesare analogă (pentru om și maimuțe) a informației vizuale în timpul activării reprezentărilor vizuale abstracte. Acestea implică reprezentări vizuale superordonate (de rang înalt), similare, ca și utilizarea implicită a asocierilor experimentate, între obiect și context.

⁵ Simon Thorpe; Denis Fize; Catherine Marlot, *Nature*, „Speed of processing in the human visual system”, *Nature*, 1996 Jun 6; 381(6582), pp. 520-522.

⁶ Denis Fize, Maxime Cauchoux, and Michèle Fabre-Thorpe, *Humans and monkeys share visual representations*, *Psychological and Cognitive Sciences*, PNAS, Proceedings of the National Academy of Sciences, U S A, 2011 May 3, 108(18), pp. 7635-7640.

⁷ Arnaud Delorme, Ghislaine Richard, Michele Fabre-Thorpe, *Rapid Processing of Complex Natural Scenes: A Role for the Magnocellular Visual Pathways?*, *Neurocomputing*, 26-27, 663-670, 1999.

Alături de formă și textură, culoarea este o caracteristică importantă pentru imagine. În testele făcute comparativ⁸ (pe primate și om), se constată că viteza de reacție la apariția unui animal într-o imagine este foarte scurtă (de la 20 ms până la 150 ms) și depinde de culoare doar când este afectată claritatea imaginii, caz în care este necesar un timp de procesare suplimentar. Se pare că tendința generală în testele realizate, relevă importanța preponderentă a culorii mai ales în testele în care trebuie denumit obiectul, mai mult decât în testele în care trebuie realizată o simplă categorizare vizuală. Culoarea intervine în categorizare doar în conceperea răspunsurilor tardive. Luarea în calcul a indicilor cromatici pare a fi lentă, această presupunere fiind influențată și de faptul că, pe lângă indicele de culoare, în testele de denumire a obiectelor, intervine (ca proces consumator de timp) și evaluarea aspectelor semantice pe care trebuie să le rezolve sistemul cognitiv uman, pentru a da denumirea, sau răspunsul corespunzător. Ca o observație colaterală, compararea performanțelor și timpilor de reacție pentru om și pentru primate arată că timpii de răspuns medii în sarcina de detecție a alimentelor au valori aproximativ egale, iar culoarea intervine mai mult în testele de detecție de aliment decât în testele de detecție de animale⁹, (cu o explicație probabil, derivând din ontologie).

Sistemul vizual trebuie să reconstruiască structura tridimensională (3D) a unui obiect din imaginile bidimensionale (2D) de pe retină. Adâncimea definită cu ajutorul gradientilor de disparitate binoculari este codată de către neuronii din lobul IT (inferior temporal)¹⁰. Cercetări relativ recente, arată că lobul temporal inferior realizează și codificarea adâncimii definite cu ajutorul gradientilor de textură. Neuronii infero-temporali singulari s-au dovedit selectivi pentru deformările de suprafață definite fie prin textură, fie prin disparitate. Perceperea înfășurătoarei unei forme proeminente, în relief, cu textură, s-a dovedit, prin măsurători, independentă de perceperea înclinării liniilor de pe suprafața sa, de perceperea pantelor sale, ceea ce indică codificarea separată a pantei (înclinării), de anvelopa formei 3D studiate. Mai mulți neuroni sunt antrenați la recunoașterea formei 3D decât la detecția pantei.

S-au utilizat înclinări de maxim 30° pentru a evita schimbări majore care să facă imaginea de nerecunoscut. Se constată că neuronii din zona infero-temporală sunt realmente mult mai sensibili la forma tridimensională (la detecția unei proeminente) decât la detecția unghiurilor înclinărilor, a liniilor oblice, din texturi (Nguyenkim and DeAngelis, 2003)¹¹.

Studii neurofiziologice recente au demonstrat importanța interacțiunilor corticale laterale și de feedback, în procesele vizuale fundamentale care includ segmentarea texturii și efectele contrastului figură-fundal^{12,13}. Experimentele

⁸ Arnaud Delorme, *Idem*.

⁹ Arnaud Delorme, *Idem*.

¹⁰ Yan Liu, Rufin Vogels, and Guy A. Orban, *Convergence of Depth from Texture and Depth from Disparity in Macaque Inferior Temporal Cortex*, The Journal of Neuroscience, Behavioral/Systems/Cognitive, 24(15):3795–3800, April/14/2004.

¹¹ Jerry D. Nguyenkim, Gregory C. DeAngelis (2003) *Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons*, Journal of Neuroscience 23:7117–7128.

¹² Jean-Michel Hupé, Andrew C. James, Bertram R. Payne, Stephen G. Lomber, Pascal Girard, & Jean Bullier, *Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons*, Nature, 394, 784-787, 1998.

neurofiziologice efectuate referitor la segmentarea texturii și detecția efectelor figură-background relevă faptul că acesta este un proces consumator de timp, care implică circuite feed-back (Lamme 1998)¹⁴, ca atare este dificil de imaginat cum un astfel de proces complex ar putea fi realizat doar cu ajutorul unei rețele neuronale feed-forward. Rezultatele referitoare la timpul efectiv de reacție în recunoașterea unor ținte percepute vizual (cum este detecția unui animal într-o scenă), sugerează că această identificare este posibil să se efectueze fără procese de detecție a detaliilor, care sunt consumatoare de timp. Se poate deduce că identificarea unui animal într-o scenă naturală poate fi făcută printr-o căutare paralelă a unui număr mare de trăsături componente. Dar, pentru creier, uneori, un singur indiciu este suficient pentru a identifica posibila prezență a unui animal într-un cadru, o parte din corp (ochi, dinte, membre) fiind suficientă pentru a declanșa un răspuns de identificare a prezenței chiar și în condițiile în care imaginea animalului (care poate fi parțial obturată), nu a fost încă segmentată. Deducem astfel că un sistem ierarhic de decizie este acționat și că acest sistem dă prioritate acestor sub-module care pot declanșa o reacție mecanică.

Rolul primordial îl are în această perceptivă, propagarea feed-forward, rolul buclilor de feedback fiind secundar, așa cum rezultă din calculul timpului de reacție extrem de scurt^{15, 16}. Acest raționament a condus la concepția tehnologiei SpikeNet coordonată de către Simon Thorpe, care utilizează o rețea neurală specială codată prin spike-uri caracterizate în intensitate și poziție temporală, ținând cont de apariția lor în timp. Fiind o rețea simplă, de recunoaștere a unui obiect (model sau pattern), fără feedback, acest nou tip de rețea neuronală dă răspunsuri suficient de bune și rapide în recunoașterea formelor obiectelor învățate din mediul înconjurător.

Deci, în ierarhia de procesare a informațiilor remarcăm faptul că forma este sesizată în primul rând, iar culoarea este utilizată în cazul în care sunt necesare detalii suplimentare pentru recunoaștere și decizie. Abia ulterior sunt observate, remarcate, analizate, zonele cu diverse texturi, din imagine, care induc imediat scenariul asociat informației respective. De exemplu dacă se identifică o textură de cărămidă, este posibil să avem un zid, dacă identificăm scoarță, este plauzibil să existe un copac, sau mai mulți, dacă identificăm iarbă poate avem un câmp sau o grădină, sau un teren de fotbal. Indicii suplimentare completează scenariul posibil identificat.

2. Aplicații posibile și importanța detecției texturii

O posibilă aplicație de computer vision pentru librării digitale este utilizarea texturii pentru adnotări și pentru descrierea conținutului.

¹³ Victor Lamme, Hans Supér, Henk Spekreijse, *Feedforward, horizontal, and feedback processing in the visual cortex*. Current Opinion on Neurobiology, 8, 529-535, 1998.

¹⁴ Lamme, *Idem*.

¹⁵ Michele Fabre-Thorpe, Arnaud Delorme, Catherine Marlot et Simon Thorpe, *A limit to the speed of processing in UltraRapid Visual Categorization of novel natural scenes*. Journal of Cognitive Neuroscience, 13, 171-180, 2000.

¹⁶ Michele Fabre-Thorpe, Ghislaine Richard et Simon Thorpe, *Rapid categorization of natural images by rhesus monkeys*, Neuroreport, 9, 303-308. (1998).

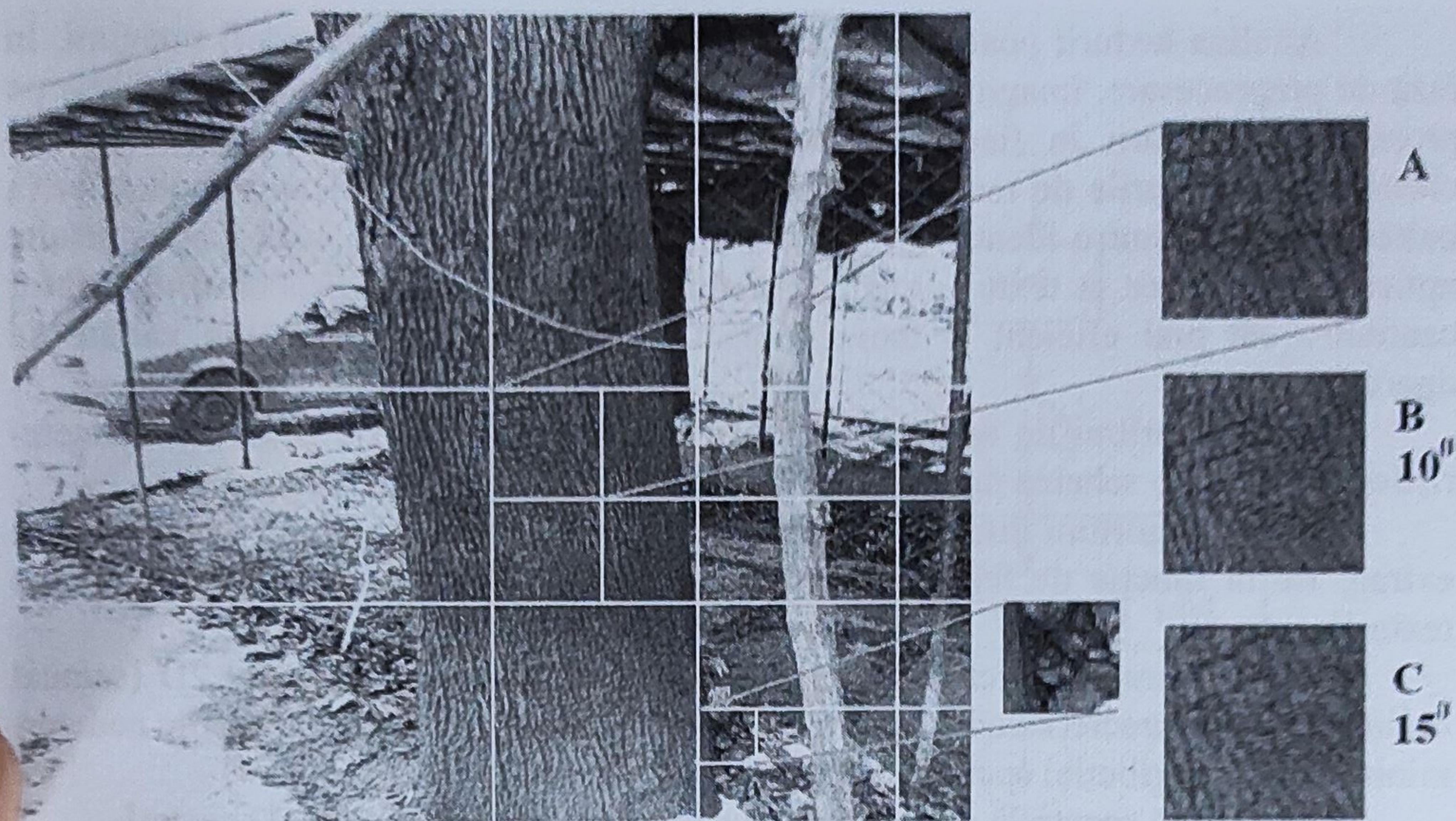


Fig. 1. *Ilustrarea cerințelor unui sistem de adnotare automată a texturii*

Adnotarea bazată pe procesări proprii domeniului vederii artificiale poate fi semi-supervizată de către utilizator în atașarea descriptorilor pentru baze extinse de date (imagini sau video-uri). Odată etichetată o anumită textură cu termenul corespunzător, modelul de textură poate fi utilizat pentru a propaga această etichetă și la alte regiuni cu o textură „vizual similară”.

Deocamdată nu există implementări care să dea rezultate suficient de bune încât să se potrivească realmente cu modul în care percepția umană detectează similaritatea în imagini. Ca urmare, pentru îmbunătățirea acurateții, pot fi selectate câteva texturi-model pentru aceeași categorie de textură, cu condiția ca aceste modele să fie bine alese și reprezentative.

Un sistem va baleia imaginea și pentru fiecare eșantion de imagine va găsi modelul de textură care se potrivește în proporția cea mai mare cu zona selectată din imagine. În cazul în care o singură textură nu corespunde suficient de bine zonei alese, fie se sub-eșantionează zona din imagine și se caută texturile corespunzătoare, fie se atribuie o nouă categorie texturii.

3. Sistem de procesare automată

Un sistem tipic de procesare de imagini (incluzând și procesarea texturii) poate parcurge mai multe etape: achiziție de imagini, preprocesare, extragere de trăsături, clasificare, postprocesare, decizie.

Referindu-ne la textură putem vorbi despre metodele de analiză a texturii, clasificarea texturii, segmentarea pe bază de textură, extragerea de forme în funcție de textura înglobată și de sinteza de texturi plecând de la eșantioane de textură.

Analiza texturii poate fi realizată în diverse etape ale procesării imaginii. În faza de preprocesare, imaginile pot fi segmentate în regiuni contigue, bazându-ne pe proprietățile texturii în fiecare regiune. În faza de extragere de trăsături și de clasificare, trăsăturile de textură pot furniza indicii pentru clasificarea modelelor (a patter-urilor) și pentru identificarea obiectelor. Element de bază între toate celelalte aplicații referitoare la texturi, analiza texturilor este menită găsirii unui descriptor cantitativ cât mai eficient și mai compact, prin utilizarea unei game variate de operatori.

Unii algoritmi de analiză a texturii implică extragerea trăsăturilor caracteristice și propun o schemă de codare care să înglobeze aceste caracteristici.

Acești algoritmi diferă fie în funcție de tipul de trăsături de textură pe care le extrag, fie în funcție de felul în care aceste trăsături sunt prezentate în descrierea texturii.

O abordare statistică descrie o textură prin statistici de semnal 2D (semnal bidimensional, caracterizând imaginea respectivă), care reflectă proprietățile nedeterministice ale distribuției spațiale ale semnalelor din imagine.

O metodă spectrală extrage trăsăturile de textură din domeniul spectral. O abordare structurală a texturii consideră textura ca o ierarhie de aranjamente spațiale de primitive de textură bine-definite.

Un model de probabilitate poate descrie procesul stohastic care stă la bază generării texturilor.

4. Procesarea texturii cu DTCWT

Vom prezenta în acest capitol câteva rezultate practice ale aplicării transformatei wavelet într-o formă propusă de Nick Kingsbury și Ivan Selesnick¹⁷ care evidențiază orientările unghiulare într-o textură și punctele de amplitudine maximă (intensitate preponderentă) care rezultă din aplicarea acestei metode, denumită Dual Tree Complex Wavelet Transform (abreviat DTCWT Transformata Wavelet Discretă cu arbore dublu, complex, în care *C* italic face diferența față de alte abrevieri posibile, uzuale – de exemplu pentru continuous wavelet transform CWT – transformata wavelet continuă).

DTCWT calculează transformata complexă a unui semnal utilizând două descompuneri separate arborele *a* și arborele *b*, create din descompuneri DWT (Discrete Wavelet Transform). Dacă filtrele utilizate în unul din acești arbori sunt proiectate diferit față de filtrele utilizate în cel de al doilea arbore, atunci este posibil ca unei DWT să îi fie atribuiți coeficienții reali iar celeilalte DWT să îi fie atribuiți coeficienții imaginari.

¹⁷ Ivan W. Selesnick, Richard G. Baraniuk, and Nick G. Kingsbury, *The Dual-Tree Complex Wavelet Transform*, IEEE Sign. Proc. Mag., vol. 22, no 6, pp 123-151, Nov. 2005.

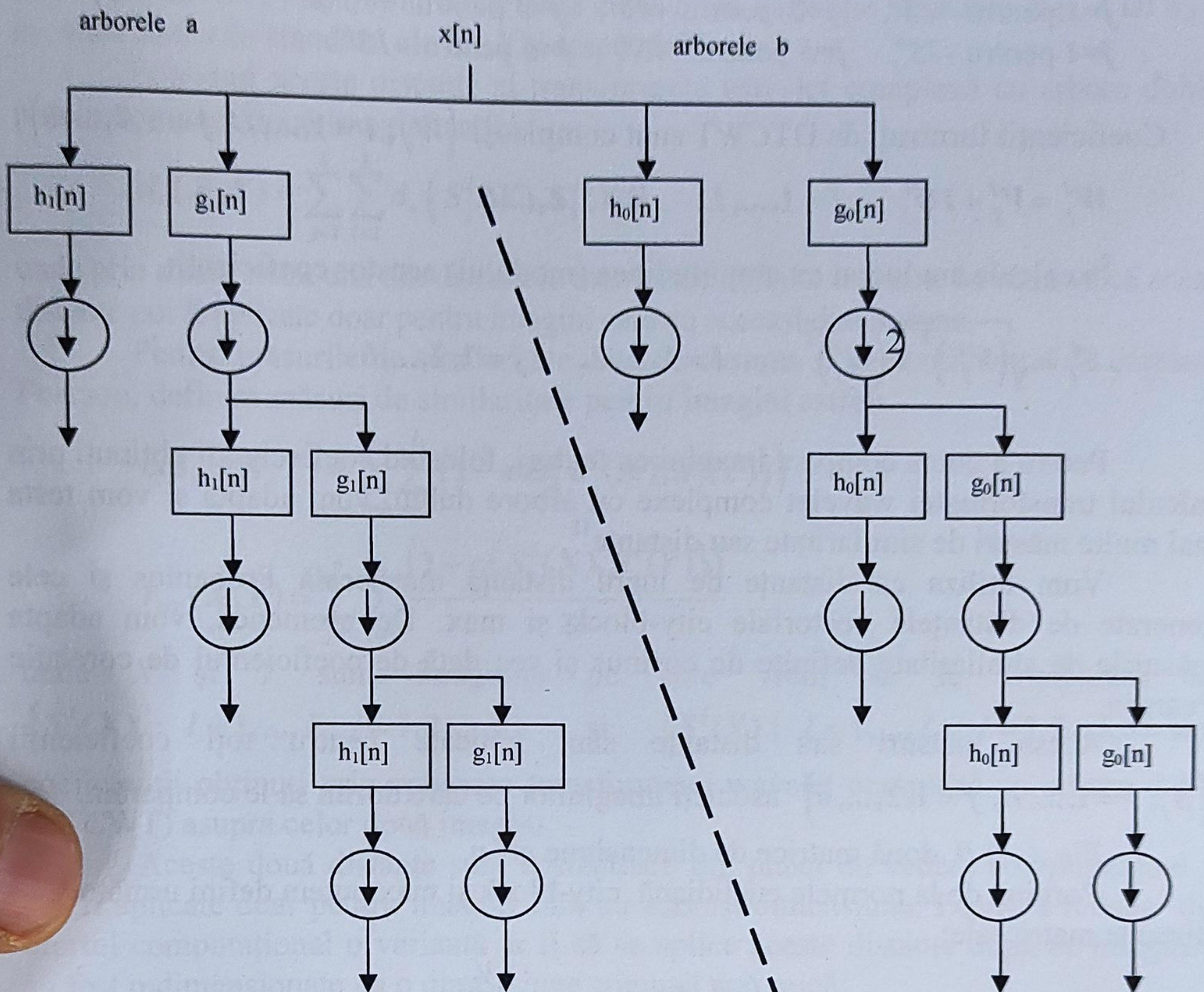


Fig. 2. DTCWT – o metodă robustă pentru determinarea trăsăturilor de textură

Transformata wavelet complexă cu acest arbore dublu furnizează 6 subimagini cu informații direcționale la $\pm 15^\circ, \pm 45^\circ, \pm 75^\circ$.

Se obțin astfel 6 seturi de coeficienți complecși

$$\{W_j^l; l = 1, \dots, L\}, \quad j = 1, 2, \dots, 6,$$

unde W_j^l sunt matrice pătratice ale căror dimensiuni se înjumătățesc pe măsură ce indexul l crește.

De exemplu, pentru o imagine de dimensiune 128×128 și $L=4$, coeficienții W_j^1 au dimensiunea 64×64 , W_j^2 au dimensiunea 32×32 , W_j^3 au dimensiunea 16×16 și W_j^4 au dimensiunea 8×8 .

Am notat cu L numărul de nivele de descompunere cu care se calculează transformata wavelet.

În calcule am folosit $L=4$.

Legătura dintre indicii j cu care se indexează subimaginele direcționale și direcții este următoarea:

$j=1$ pentru $+15^\circ$, $j=2$ pentru $+45^\circ$, $j=3$ pentru $+75^\circ$,
 $j=4$ pentru -75° , $j=5$ pentru -45° , $j=6$ pentru -15° .

Coeficienții furnizați de DTCWT sunt complecși $\{W_j^l; l=1, \dots, L, j=1, 2, \dots, 6\}$

$$W_j^l = V_j^l + iU_j^l \quad l=1, \dots, L, \quad j=1, 2, \dots, 6$$

În calcule am lucrat cu amplitudinea (modulul) acestor coeficienți:

$$S_j^l = \sqrt{(V_j^l)^2 + (U_j^l)^2} \quad , \quad l=1, \dots, L, \quad j=1, 2, \dots, 6$$

Pentru a putea compara imagini cu texturi, folosind coeficienții obținuți prin calculul transformatei wavelet complexe cu arbore dublu, vom adapta și vom testa mai multe măsuri de similaritate sau distanțe¹⁸.

Vom utiliza ca distanțe de lucru distanța matriceală Frobenius și cele generate de distanțele vectoriale city-block și max. De asemenea, vom adapta măsurile de similaritate definite de cosinus și cea dată de coeficientul de corelație Pearson.

Aceste măsuri sau distanțe sunt aplicate pentru toți coeficienții $\{S_j^l; l=1, \dots, L, j=1, 2, \dots, 6\}$ asociați imaginilor pe care dorim să le comparăm.

Fie A și B două matrice de dimensiune $m \times n$.

Pornind de la normele euclidiană, city-block și max putem defini următoarele distanțe matriceale:

$$d_2(A, B) = d_{fro}(A, B) = \left(\sum_{i=1}^m \sum_{j=1}^n (a_{ij} - b_{ij})^2 \right)^{1/2}$$

$$d_1(A, B) = \max \left\{ \sum_{i=1}^m |a_{ij} - b_{ij}|; \quad j=1, 2, \dots, n \right\}$$

$$d_\infty(A, B) = \max \left\{ \sum_{j=1}^n |a_{ij} - b_{ij}|; \quad i=1, 2, \dots, m \right\}$$

Fie x și y doi vectori n -dimensionali.

Măsura de similaritate cosinus pentru vectorii x și y este dată de relația:

$$\cos(x, y) = \frac{y^T x}{\|x\| \|y\|}$$

unde prin $\|\cdot\|$ am notat norma euclidiană.

Coeficientul de corelație Pearson pentru două variabile aleatoare X și Y se definește ca fiind:

$$r(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

¹⁸ Rui Xu, Donald C. Wunsch II, *Clustering*, IEEE Press/Wiley 2009.

unde prin $\text{cov}(X, Y)$ am notat covarianța celor două variabile aleatoare X și Y iar σ_X și σ_Y sunt abaterile standard ale lui X și respectiv Y .

Folosind aceste distanțe și transformata wavelet complexă cu arbore dublu, putem compara două imagini astfel:

$$d_*(X, Y) = \sum_{j=1}^6 \sum_{l=1}^L d_*(S'_j(X), S'_j(Y))$$

unde prin d_* am notat una din distanțele matriceale definite mai sus. Se observă că aceste distanțe pot fi aplicate doar pentru imagini care au aceeași dimensiune.

Pentru măsurile de similaritate date de cosinus și de coeficientul de corelație Pearson, definim măsuri de similaritate pentru imagini astfel:

$$\cos_{DT}(X, Y) = \sum_{j=1}^6 \sum_{l=1}^L \left(1 - \cos(S'_j(X), S'_j(Y)) \right)$$

$$r_{DT}(X, Y) = \sum_{j=1}^6 \sum_{l=1}^L \frac{\left(1 - r(S'_j(X), S'_j(Y)) \right)}{2}$$

unde X și Y sunt imaginile pe care vrem să le comparăm iar $\{S'_j(X); l=1, \dots, L, j=1, 2, \dots, 6\}$ și $\{S'_j(Y); l=1, \dots, L, j=1, 2, \dots, 6\}$ sunt coeficienții obținuți prin aplicarea transformatei wavelet complexă cu arbore dublu (DTCWT) asupra celor două imagini.

Aceste două distanțe sunt costisitoare din punct de vedere computațional și pot fi aplicate doar pentru imagini care au aceeași dimensiune. Pentru a reduce din efortul computațional o variantă ar fi să se aplice aceste distanțe după ce imaginile au fost redimensionate la o dimensiune comună mai mică:

$$\cos_{DT}^{ec}(X, Y) = \cos_{DT}(\text{resize}(X, \text{new_dim}), \text{resize}(Y, \text{new_dim})).$$

Această nouă definiție are avantajul suplimentar că aceste distanțe pot fi aplicate unor imagini de dimensiuni diferite.

Pentru a aplica măsura de similaritate cosinus definită mai sus pentru coeficienții matriceali considerați (cei obținuți prin aplicarea transformatei wavelet complexe cu arbore dublu), am linearizat matricele respective în ordinea liniilor.

Pentru a testa eficacitatea distanțelor/măsurilor de asemănare a texturilor construite utilizând transformata wavelet complexă cu arbore dublu (DTCWT), am folosit, în prima etapă, o bază de date de imagini împărțită pe clase de texturi¹⁹. Toate imaginile din bazele de date folosite sunt alb-negru și au dimensiunea 128×128 și au fost scalate la intervalul $[0, 1]$. Software-ul folosit în calcule care implementează transformata wavelet cu arbore dublu a fost descărcat de pe pagina Prof. Nick Kingsbury²⁰

Prima bază de date utilizată, notată cu T , conține 12 clase de texturi: scoartă, cărămizi, perete, sticlă, blană, pietre, catifea, gresie, apă, covor, lemn și pled. În fiecare din aceste clase avem câte 40 de imagini de texturi. Cele 40 de elemente din

¹⁹ Svetlana Lazebnik, Cordelia Schmid, Jean Ponce, *A Sparse Texture Representation Using Local Affine Regions*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1265-1278, august 2005.

²⁰ <http://www-sigproc.eng.cam.ac.uk/~ngk/>

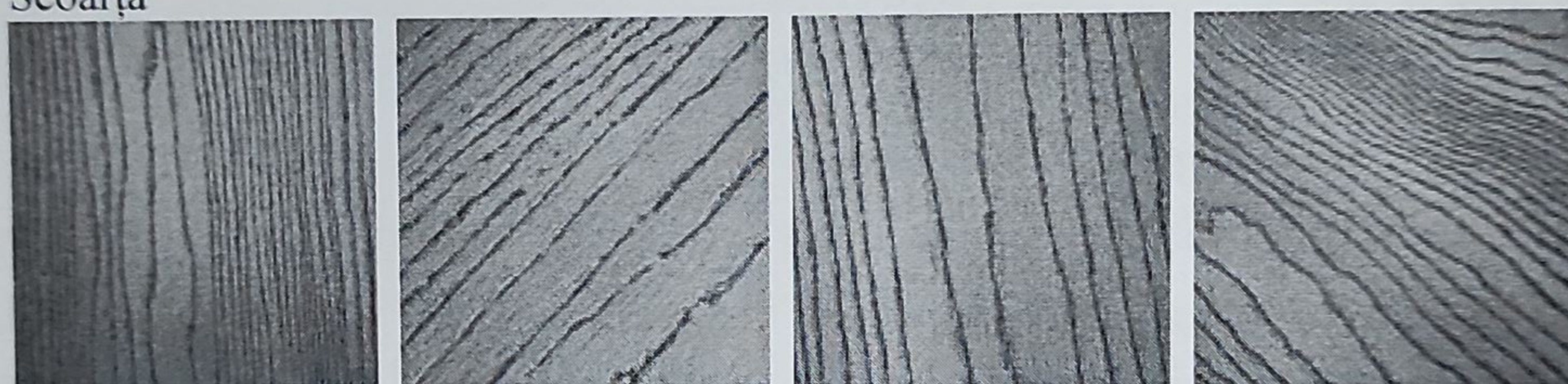
MIHAELA COSTIN, ANCA IGNAT

fiecare clasă sunt imagini ale aceleași texturi care a fost fotografiată din unghiuri diferite și la scale diferite, de asemenea la fotografiere au fost folosite diverse tipuri de iluminare.

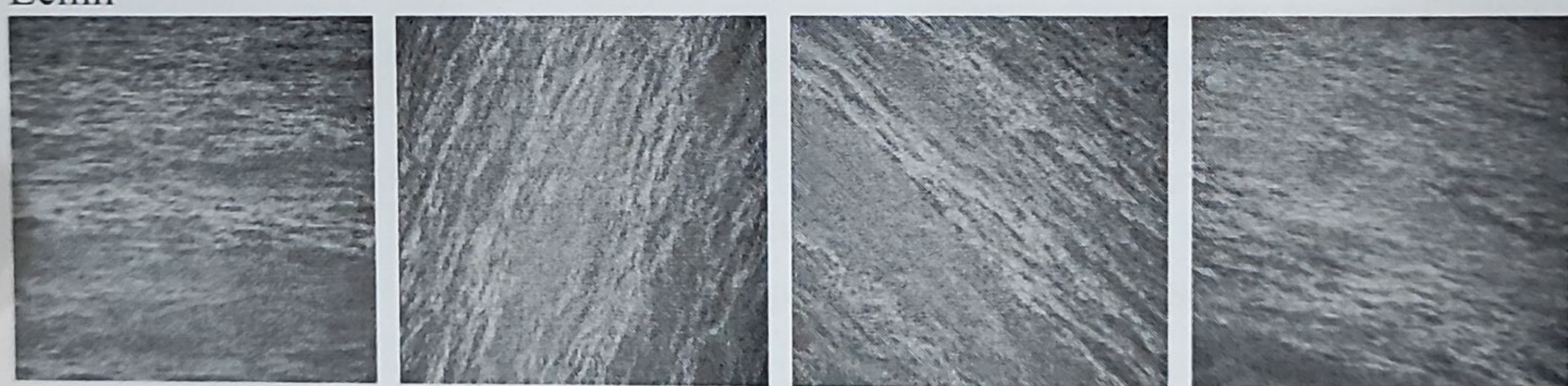
Figura 1 prezintă, spre exemplificare, câte 4 imagini din fiecare clasă.



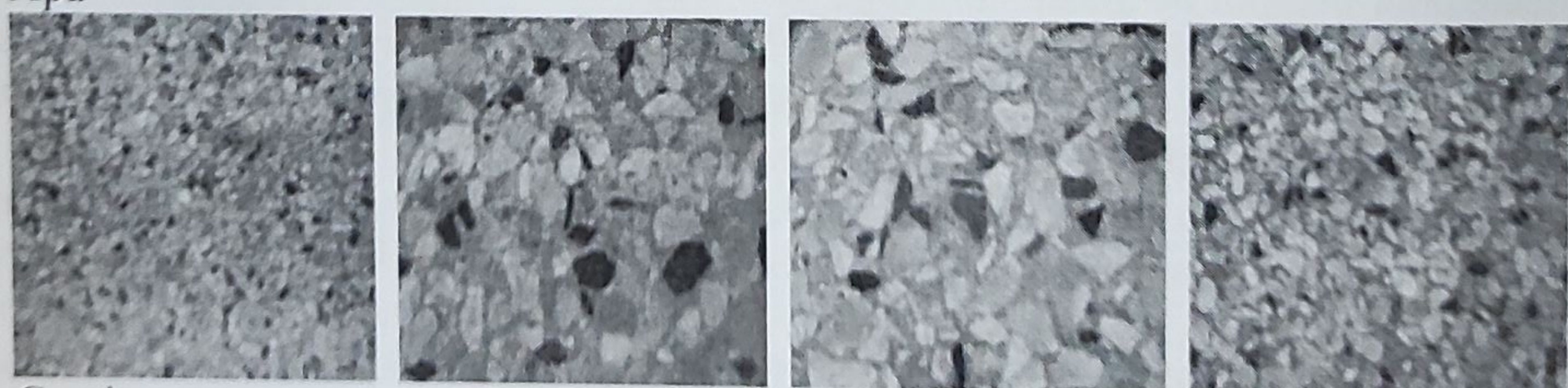
Scoată



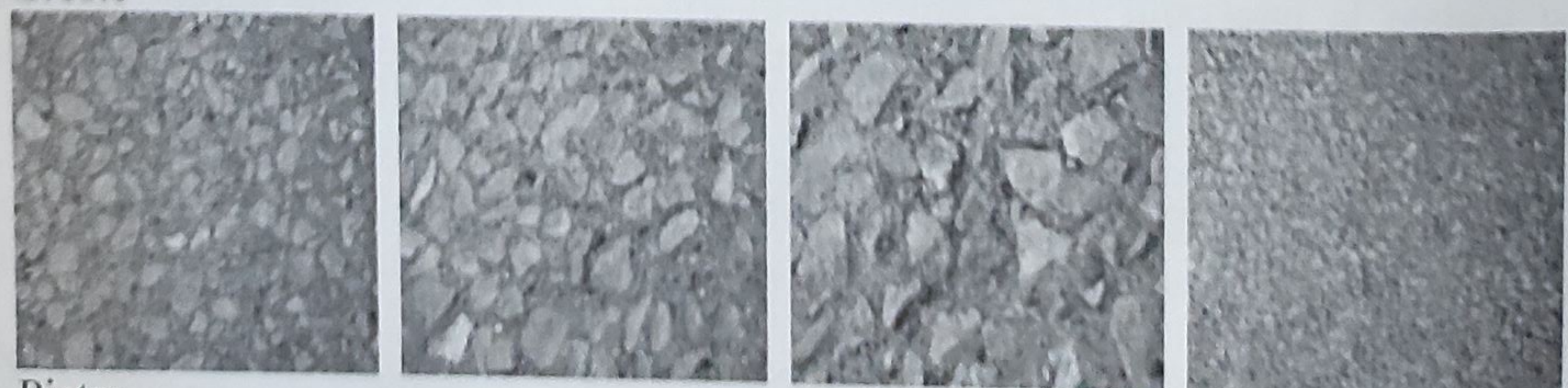
Lemn



Apă

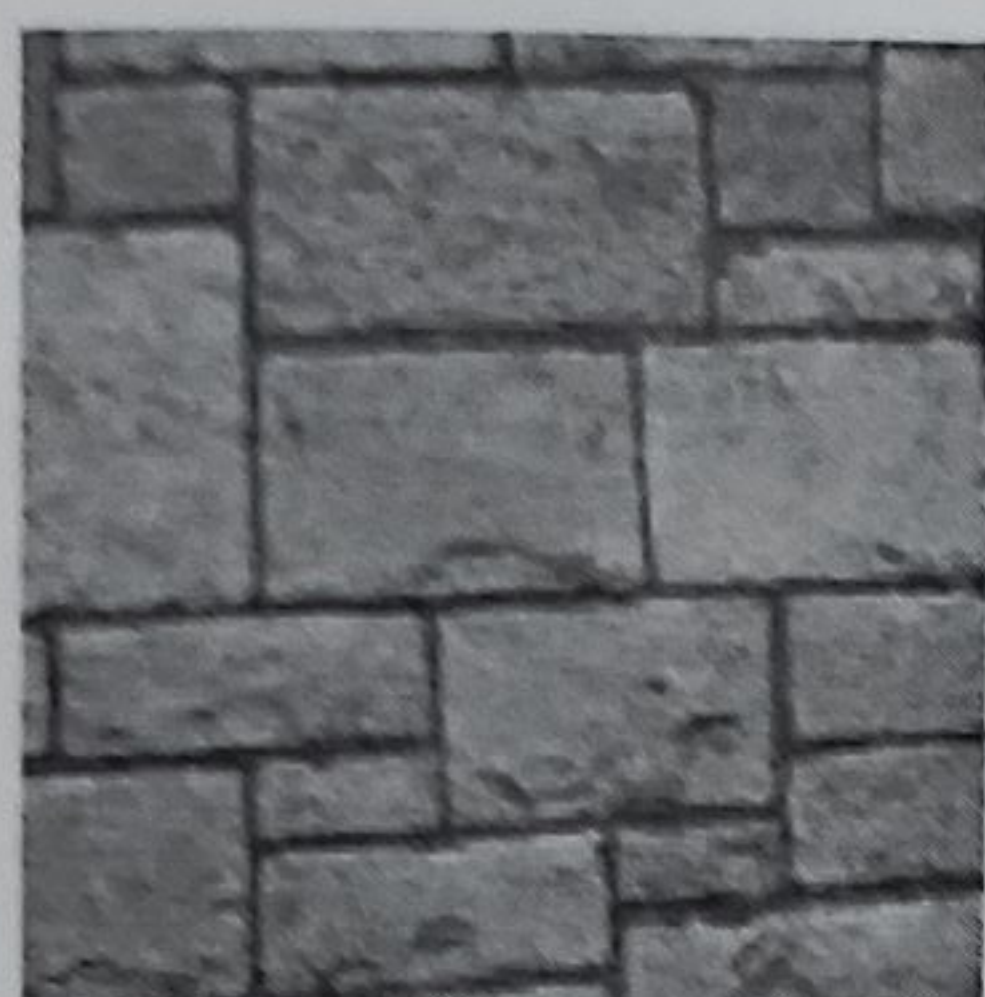


Gresie

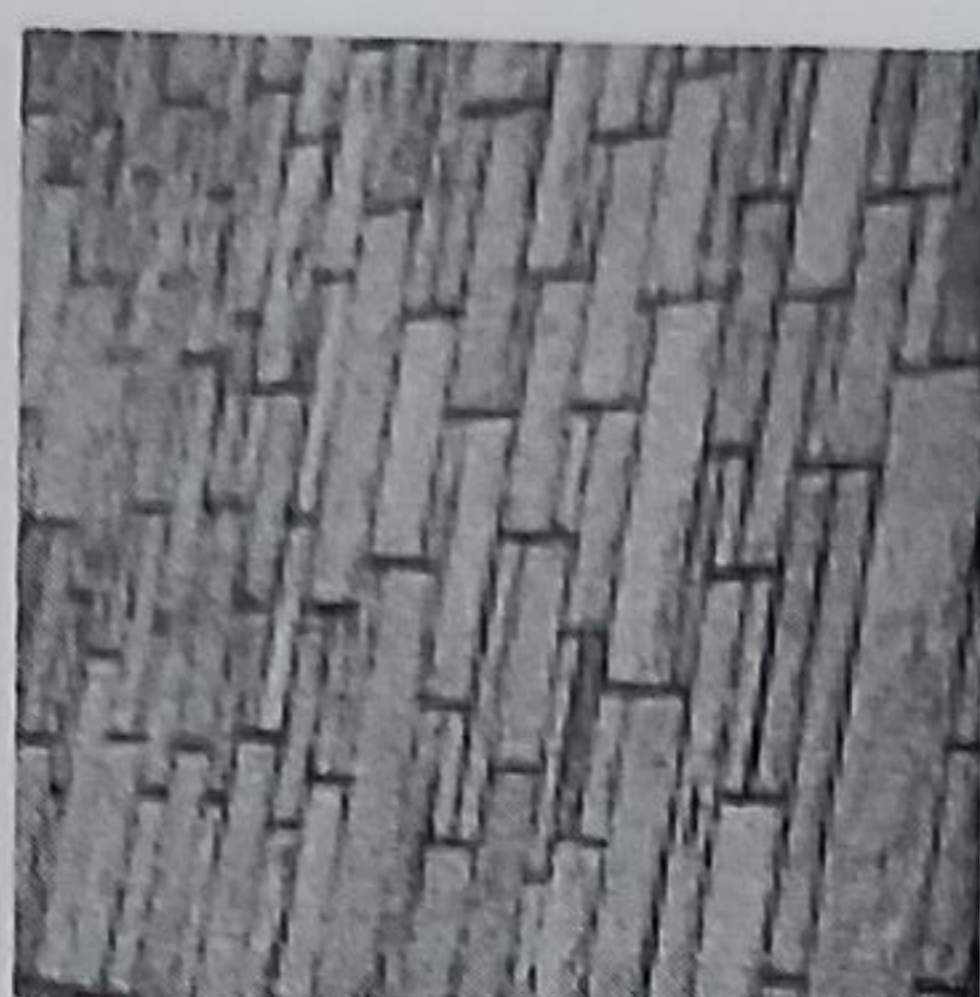


Pietre

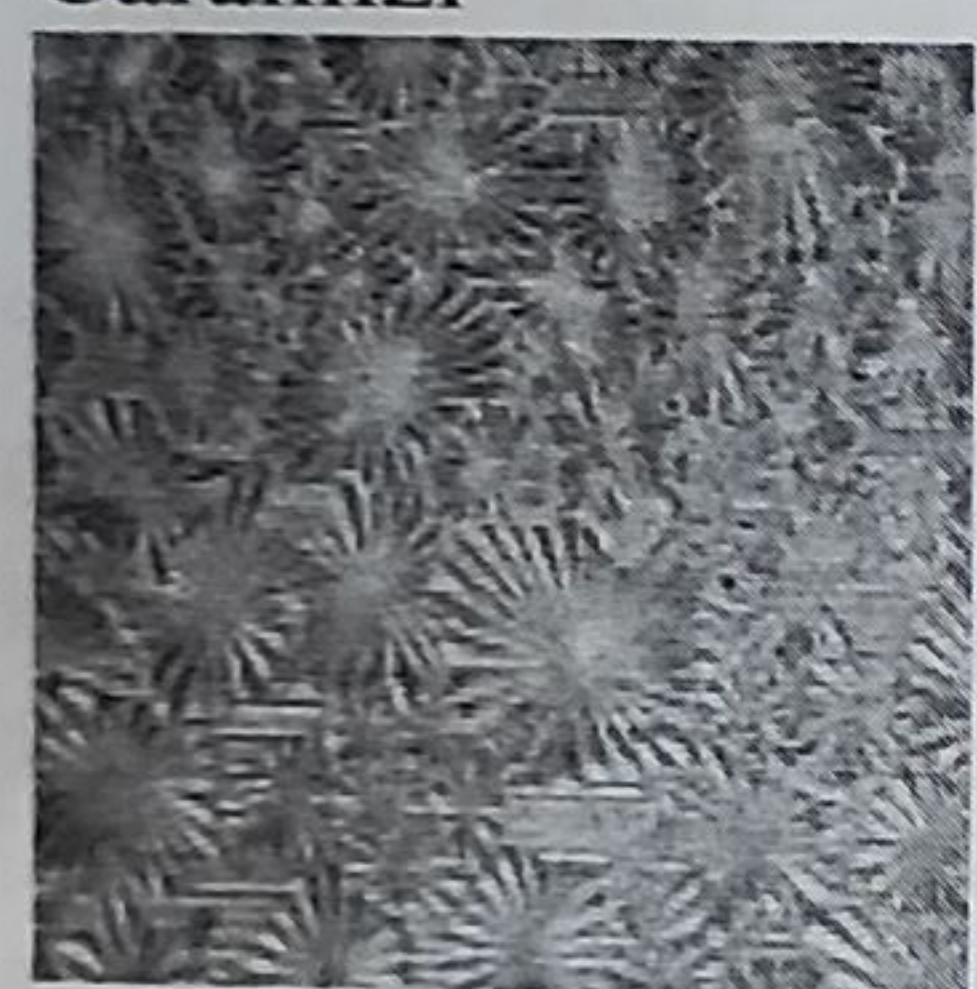
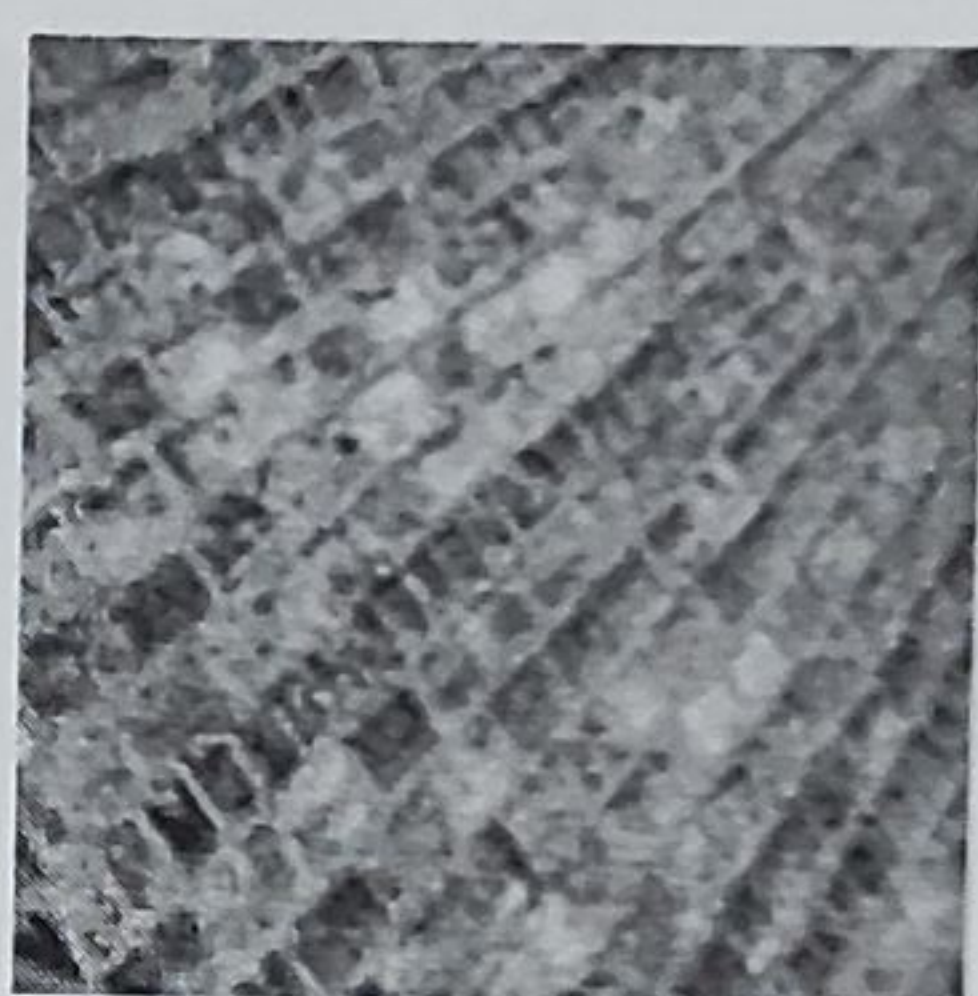
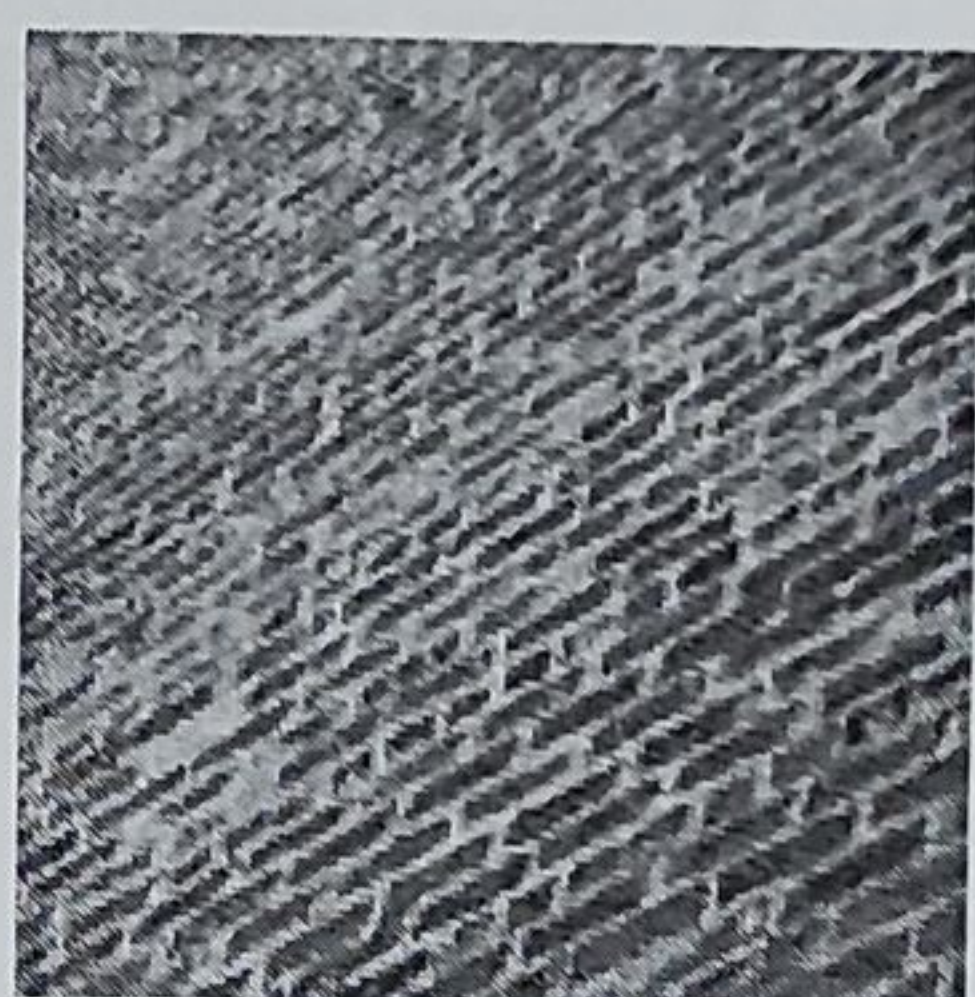
De la percepția vizuală către detecția automată a texturii



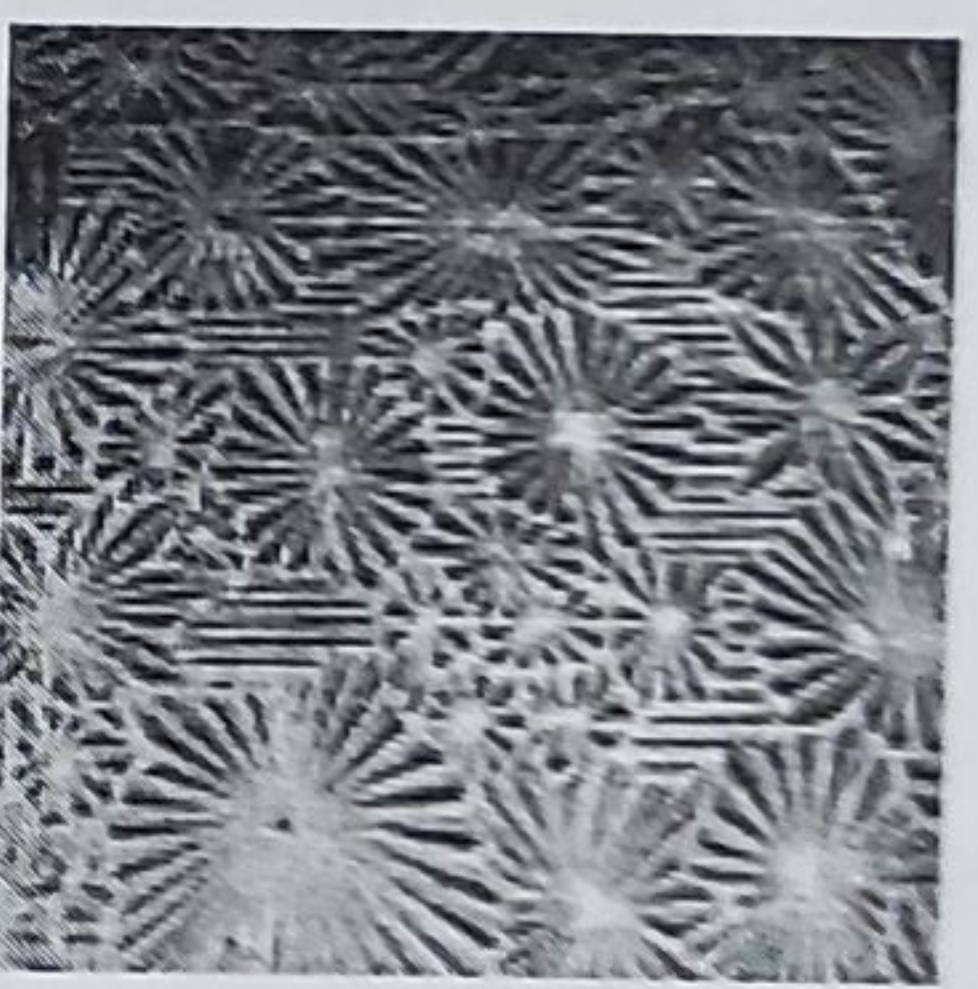
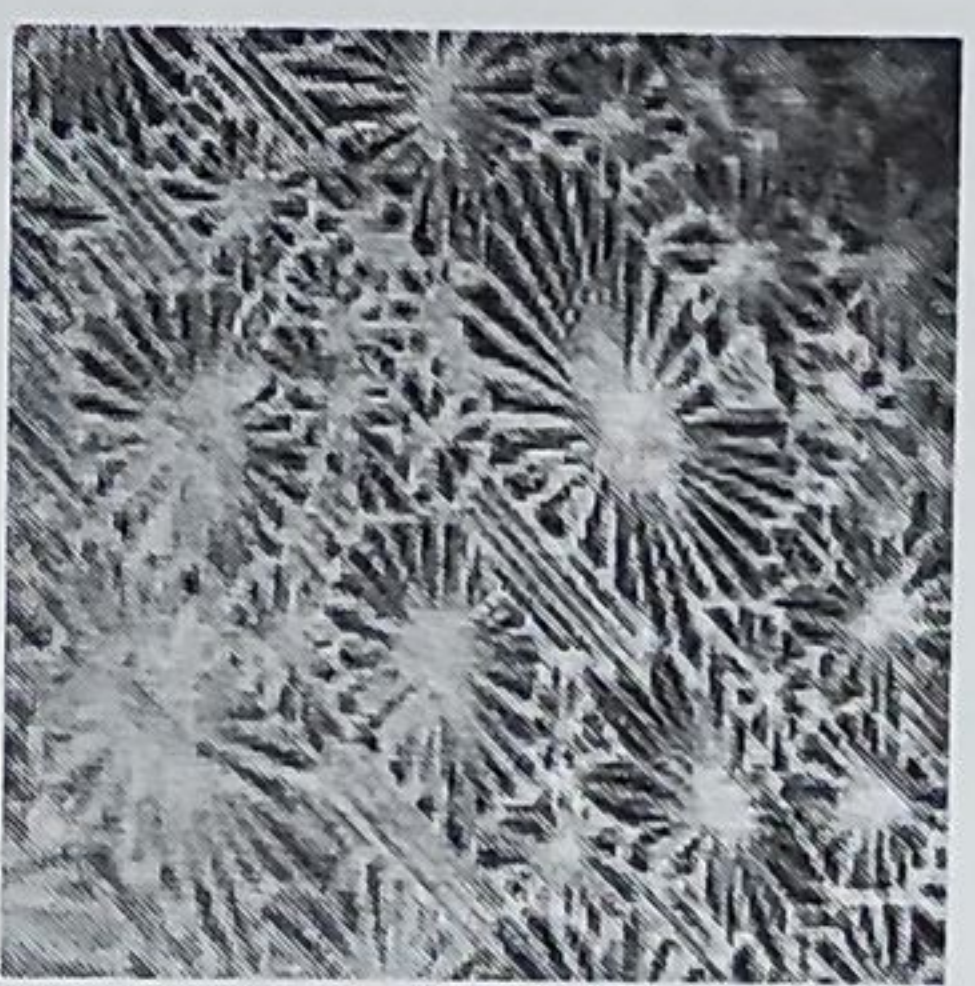
Perete



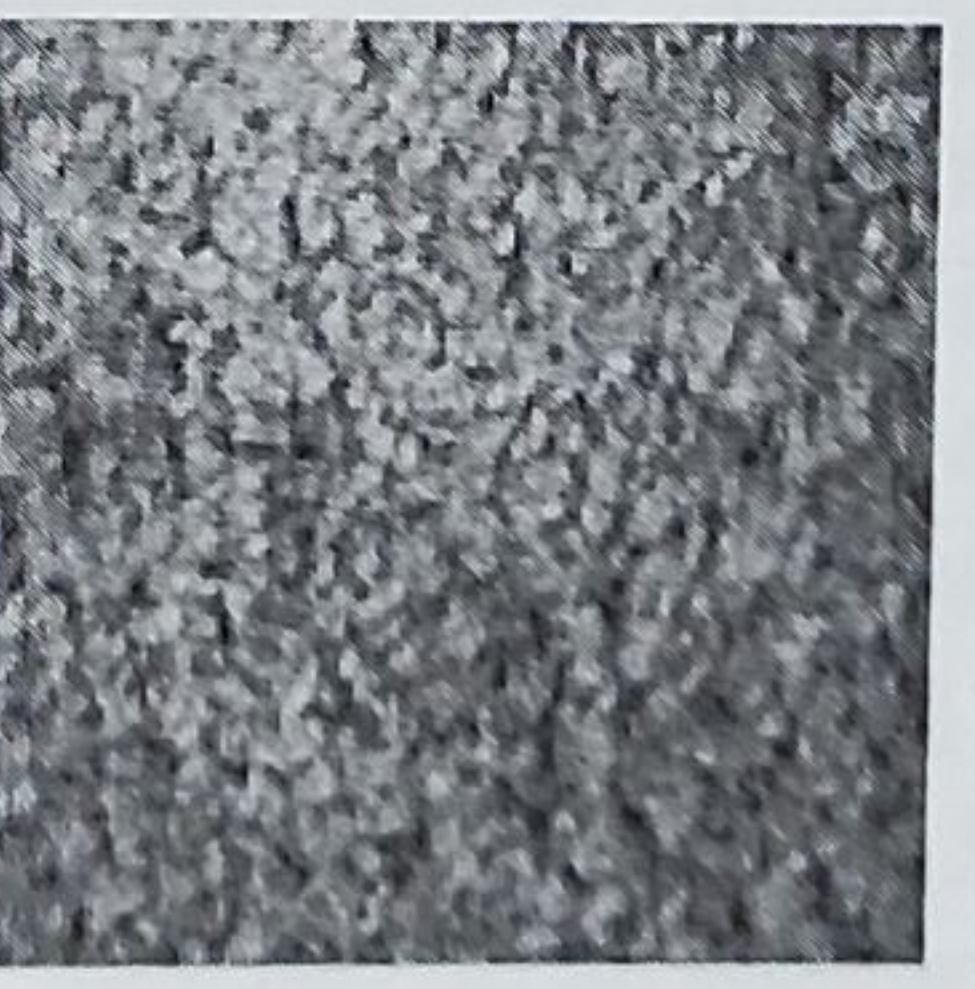
Cărămizi



Sticlă



Covor



Blană



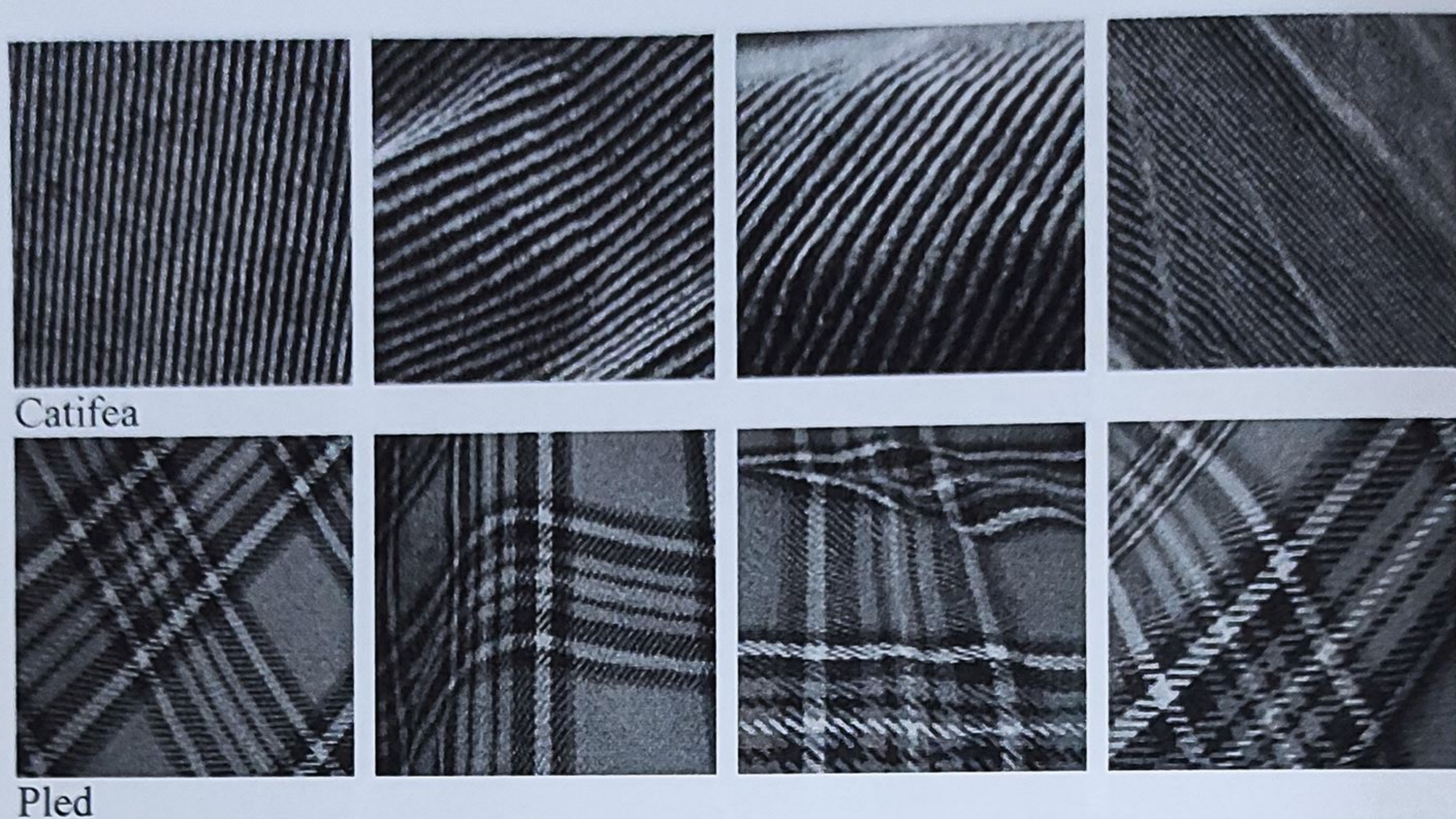


Fig. 3. Baza de date de texturi T

Pentru a testa cele cinci distanțe/măsuri de similaritate introduse, am rotit cu 10° și 15° prima imagine din fiecare clasă și apoi am aplicat aceste distanțe pentru a compara imaginile rotite cu cele din baza de date. Pentru fiecare imagine rotită, am calculat distanțele față de toate imaginile din baza de date, apoi am ordonat crescător aceste distanțe. În tabelele de mai jos am trecut pentru fiecare imagine rotită clasele a trei imagini din baza de date, acele imagini/clase care au furnizat distanțele cele mai mici față de imaginea rotită considerată.

În Tabelul I sunt trecute rezultatele acestor comparații folosind cele cinci distanțe/măsuri de similaritate propuse pentru imaginile rotite cu 10° iar în Tabelul II sunt rezultatele acestor calcule pentru imaginile rotite cu 15° .

Pentru imaginile rotite cu 10° distanțele/măsurile de similaritate propuse au nimerit clasa din care provenea imaginea rotită astfel: cu d_2 și d_∞ s-au obținut șapte rezultate corecte, cu d_1 s-au obținut cinci rezultate corecte, cu \cos_{DT} și cu coeficientul de corelație Pearson r_{DT} doar trei rezultate corecte.

Pentru imaginile rotite cu 15° avem următoarele rezultate: d_1 , d_2 și d_∞ produc șase clase exacte, cu coeficientul de corelație Pearson r_{DT} se obțin rezultate corecte pentru patru texturi iar cu \cos_{DT} două clase corecte.

Din cele două tabele observăm că distanțele d_1 , d_2 și d_∞ tind să furnizeze, în marea majoritate a cazurilor, rezultate asemănătoare.

Texturile pentru care se obțin rezultatele cele mai bune sunt: scoarța, lemnul, apa și catifeaua. Texturile care nu au fost aproape niciodată identificate sunt gresia, peretele, sticla și covorul, dar, privind aceste imagini rezultatul este justificat, explicabil, întrucât imaginile sunt dificil distincibile chiar și de către privirea umană, imaginile fiind ușor confundabile).

Din tabele se observă de asemenea că măsura de similaritate \cos_{DT} în foarte multe dintre cazuri plasează catifeaua pe prima poziție (adică, ar fi textura cea mai apropiată de marea majoritate a imaginilor rotite).

În fiecare clasă a bazei de date sunt imagini ale aceleiași texturi, dar imaginile sunt foarte diverse din punct de vedere al direcțiilor, proprietate a transfor-

matei wavelet complexe cu arbore dual pe care încercăm să o folosim pentru identificare/clasificarea texturilor.

Deoarece rezultatele de identificare obținute pentru imaginile rotite folosind baza de date T sunt mai mult nesatisfăcătoare, am creat o nouă bază de date (pornind de la baza de date T) în care am redus numărul imaginilor din fiecare clasă, păstrând prima imagine din baza de date și alte imagini care au, în mare, aceleași direcții ca și aceasta.

În noua bază de date, în fiecare clasă sunt între 5 și 10 imagini. Am reluat calculele cu fișierele rotite cu 10° și 15° folosind această bază de date, redusă. Rezultatele calculelor se găsesc în Tabelele IV și V.

Pentru clasarea imaginilor rotite cu 10° (Tabelul III) avem următoarele rezultate: distanța d_1 nimerește clasa de proveniență a imaginii rotite în șapte cazuri, d_2 în 8 cazuri, d_∞ în șapte situații, măsura \cos_{DT} în patru cazuri și coeficientul de corelație Pearson r_{DT} în șapte cazuri.

Pentru imaginile rotite cu 15° (Tabelul IV) avem următoarele rezultate: distanța d_1 găsește rezultatul corect în șase cazuri, d_2 și d_∞ în șapte situații, măsura \cos_{DT} în două cazuri și coeficientul de corelație Pearson r_{DT} în cinci cazuri. Rezultatele sunt mai bune decât cele pentru baza de date completă. Ca și în cazul precedent, texturile scoarță, lemn, apă și pietrele, sunt texturile unanim recunoscute de distanțele d_1 , d_2 și d_∞ . Texturile care nu sunt identificate sunt, ca în situația precedentă, gresia, peretele, cărămizile, sticla și covorul.

Chiar și cu această modificare a bazei de date, rezultatele obținute nu sunt mulțumitor de performante. Pentru a vedea dacă rezultatele pot fi îmbunătățite și a studia influența componenței claselor asupra rezultatelor, am redus baza de date astfel încât în fiecare clasă să fie o singură imagine și anume imaginea din care s-au construit imaginile rotite. Am făcut, din nou, aceleași calcule, iar rezultatele se regăsesec în Tabelele V (pentru imaginile rotite cu 10°) și VI (pentru imaginile rotite cu 15°). Rezultatele nu s-au îmbunătățit față de situația comparațiilor cu baza de date redusă și nici măcar față de cele obținute când am folosit baza de date completă. În cazul imaginilor rotite cu 10° d_1 , d_2 și d_∞ produc fiecare șase rezultate corecte, \cos_{DT} patru rezultate corecte și r_{DT} șapte rezultate bune, iar pentru imaginilor rotite cu 15° d_1 furnizează cinci rezultate bune, d_2 și d_∞ produc fiecare șase rezultate corecte, \cos_{DT} trei rezultate corecte și r_{DT} opt rezultate exacte. Texturile pentru care se obțin rezultate corecte pentru toate distanțele sunt aceleași ca și în cazurile precedente.

În concluzie, indiferent de numărul de elemente din clasele bazei de date T, rezultatele obținute folosind cele cinci distanțe/măsuri de similaritate propuse pentru compararea imaginilor nu sunt performante.

În toate cele trei cazuri, texturile pentru care se obțin rezultate exacte/greșite sunt aproximativ aceleași. Din rezultatele afișate în primele șase tabele, deducem că distanța cu care se obțin cele mai multe rezultate corecte este d_2 urmată de d_1 și d_∞ iar rezultatele mai puțin performante sunt obținute cu măsura dată de cosinus, \cos_{DT} . Pentru comparație, am schimbat în continuare baza de date și am folosit albumul de imagini cu texturi Brodatz²¹²².

²¹ Albumul cu texturi Brodatz - <http://sipi.usc.edu/database/>

²² Phil Brodatz, *Textures: A Photographic Album for Artists and Designers*, Dover Publications, New York, 1966.

În baza de date Brodatz cu care am lucrat sunt 13 clase de texturi: apă, cărămizi, iarbă, lână, lemn, nisip, paie, piele, piele de porc, rafie, scoarță, spumă și țesătură. În fiecare din clasele albumului Brodatz avem exact 7 imagini: o imagine a texturii care caracterizează din punct de vedere vizual clasa respectivă și 6 variante rotite ale acestei imagini.

Unghiurile de rotație folosite pentru a obține celelalte 6 imagini sunt 30° , 60° , 90° , 120° , 150° și 200° . Prin urmare în albumul Brodatz folosit avem 91 imagini împărțite în 13 clase.

În Figura 2 se găsesc imaginile texturilor de bază din albumul Brodatz.

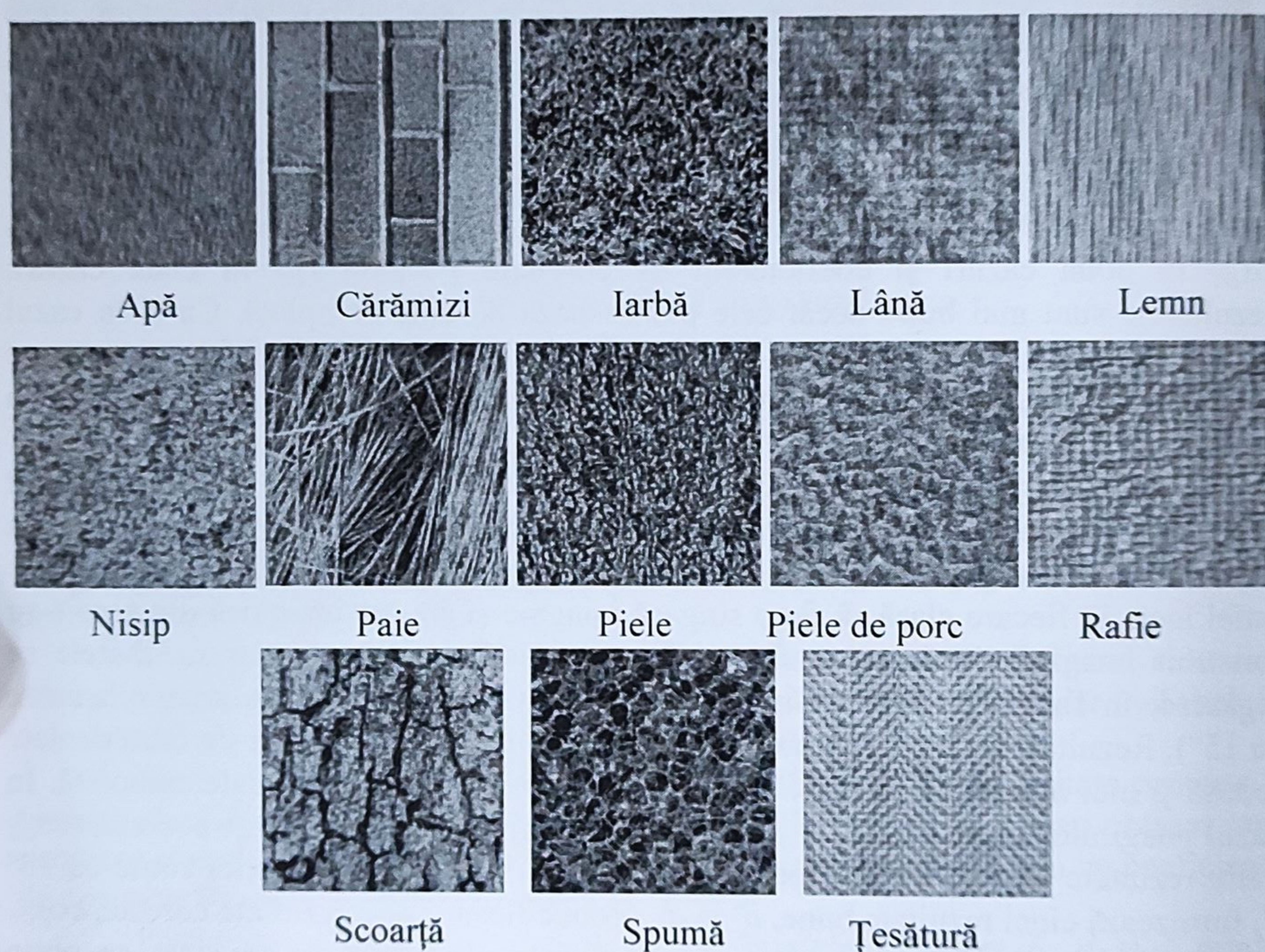


Fig. 4. Albumul de texturi Brodatz – cele 13 imagini de bază

Am reluat calculele de clasificare a imaginilor rotite cu 10° și 15° pentru noua bază de date. Rezultatele sunt în Tabelele VII (pentru imaginile rotite cu 10°) și VIII (pentru imaginile rotite cu 15°) și sunt radical diferite.

Când comparăm imagini rotite cu 10° fiecare distanță greșește o singură dată, pentru aceeași imagine: d_1 , d_2 , \cos_{DT} și r_{DT} nu identifică bine textura spumă iar d_∞ nu reușește să identifice corect cărămizile.

Pentru imaginile rotite cu 15° se fac mai multe clasificări eronate: d_1 – două identificări greșite, d_2 – trei identificări greșite, d_∞ face patru erori de identificare, \cos_{DT} – două greșeli și r_{DT} nu identifică corect o singură textură.

Am considerat și problema comparării/clasificării unor imagini care nu sunt în bazele de date folosite mai sus. Am ales patru imagini de scoarță de copac: două

dintre ele sunt din baza de imagini de texturi VisTex²³ și două sunt porțiuni din imagini ale unor scene cu copaci din Iași (Fig. 3). Imaginile din care au fost selectate ultimele două texturi din Fig. 3 se găsesc în Fig. 4.

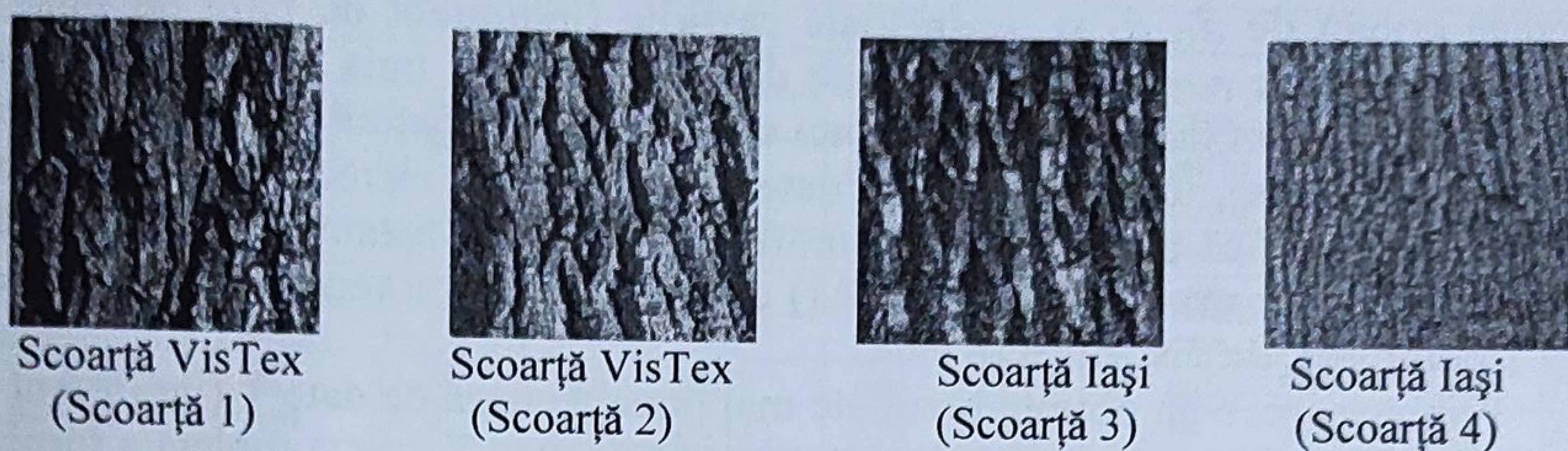


Fig. 5. Exemple - scoarțe de copac ce nu se regăsesc în bazele de date T și Brodatz



Fig. 6. Imagini din Iași cu scoarță de copac

Imaginile alese pentru a fi comparate cu elementele din cele două baze de date seamănă cu texturile din baza de date T. Ar fi normal ca să obținem rezultate mai bune pentru comparațiile cu imaginile din baza de date T. Am evaluat din nou distanțele între aceste texturi și cele din bazele de date considerate mai sus. Rezultatele se găsesc în Tabelul IX pentru comparațiile cu imaginile din baza de date completă T, în Tabelul X pentru comparațiile cu imaginile din baza de date T redusă, în Tabelul XI pentru comparațiile cu imaginile din baza de date T cu o singură textură în fiecare clasă și în Tabelul XI pentru comparațiile cu imaginile din albumul de texturi Brodatz. Prima imagine cu scoarță de copac (Scoarță 1) este identificată de distanțele d_1 , d_2 și d_x în toate cazurile când am folosit baza de date T (indiferent de variantă). Imaginea Scoarță 1 este clasată ca fiind din clasa scoarță și de distanțele

²³ Baza de date de texturi VisTex - <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/>

cosinus și coeficientul de corelație Pearson doar în cazul când s-a lucrat cu baza de date T redusă. Textura Scoarță1 nu este clasificată corect de niciuna din distanțele cu care lucrăm atunci când utilizăm albumul Brodatz. A doua textură (Scoarță2) este identificată corect de d_1 , d_2 și d_∞ în toate cazurile (indiferent de baza de imagini folosită) și în plus de r_{DT} în cazul bazei de date T redusă. A treia textură (Scoarță3) este identificată corect de d_1 , d_2 și d_∞ atunci când a fost comparată cu baza de date T completă și cea redusă. În cazul bazei de date T cu un singur element în fiecare clasă doar d_2 și d_∞ o clasifică bine. Pentru albumul Brodatz, doar distanța d_∞ reușește să o identifice corect. A patra textură (Scoarță4) este identificată corect doar pentru baza de date T completă, de distanțele d_2 și d_∞ .

Prin urmare, s-au obținut rezultate mai bune cu baza de date T (indiferent de variantă) decât cu albumul Brodatz, iar între bazele de date T, cea completă a furnizat rezultatele cele mai bune.

Am sumat toate rezultatele – clasarea celor 28 de imagini (12 rotite cu 10° , 12 rotite cu 15° și 4 texturi de scoarță) pentru baza de date T și a celor 30 de imagini (13 rotite cu 10° , 13 rotite cu 15° și 4 texturi de scoarță) și obținem următoarele procente de reușită (Tabelul XII):

- distanța d_1 : 50% pentru clasificări folosind baza de date T completă, 57,14% pentru baza de date T redusă, 46,43% pentru baza de date T cu un singur element în fiecare clasă și 80% pentru albumul de texturi Brodatz;
- distanța d_2 : 57,14% pentru clasificări folosind baza de date T completă, 60,71% pentru baza de date T redusă, 50% pentru baza de date T cu un singur element în fiecare clasă și 76,67% pentru albumul Brodatz;
- distanța d_∞ : 60,71% pentru clasificări folosind baza de date T completă, 60,71% pentru baza de date T redusă, 53,57% pentru baza de date T cu un singur element în fiecare clasă și 76,67% pentru albumul Brodatz;
- distanța \cos_{DT} : 17,86% pentru clasificări folosind baza de date T completă, 25% pentru baza de date T redusă, 28,57% pentru baza de date T cu un singur element în fiecare clasă și 73,33% pentru albumul Brodatz;
- distanța r_{DT} : 32,14% pentru clasificări folosind baza de date T completă, 46,43% pentru baza de date T redusă, 53,57% pentru baza de date T cu un singur element în fiecare clasă și 80% pentru albumul Brodatz.

Din aceste date, constatăm că cele mai bune rezultate se obțin pentru baza de date Brodatz iar dintre distanțe rezultatele cele mai bune se obțin cu d_∞ - pentru bazele de date T și d_1 și coeficientul de corelație Pearson, r_{DT} – pentru albumul Brodatz. Comparând procentele de reușită doar pentru cele trei variante de baze de date T, cele mai bune clasificări se obțin pentru baza de date T redusă.

De la percepția vizuală către detecția automată a texturii

Tabel 1. Distanțele d_1, d_2, d_∞ cosinus și corelația Pearson aplicate coeficienților dtcwt pentru prima textură din fiecare clasă rotită cu 10° comparată cu baza de date completă

	Distanțe				
	d_1	d_2	d_∞	\cos_{DT}	r_{DT}
Scoarță.10	Scoarță 63.39	Scoarță 54.64	Scoarță 58.79	Covor 5.71	Lemn 10.57
	Scoarță 65.35	Scoarță 55.41	Pietre 59.70	Catifea 5.71	Pled 10.61
	Scoarță 65.92	Căramizi 55.77	Gresie 59.93	Catifea 5.71	Pled 10.69
Lemn.10	Lemn 42.71	Lemn 32.80	Lemn 33.92	Catifea 5.61	Pled 9.86
	Lemn 43.98	Lemn 33.22	Lemn 34.09	Catifea 5.81	Pled 10.10
	Lemn 44.53	Lemn 33.38	Lemn 34.29	Lemn 5.89	Lemn 10.12
Apă.10	Apă 25.76	Apă 24.37	Apă 27.92	Catifea 5.93	Apă 10.27
	Apă 26.36	Apă 24.42	Apă 28.42	Apă 6.00	Pled 10.33
	Apă 26.40	Apă 24.73	Apă 28.84	Apă 6.00	Lemn 10.46
Gresie.10	Covor 51.13	Covor 47.16	Covor 50.70	Catifea 5.72	Sticlă 10.79
	Covor 51.70	Pietre 47.60	Pietre 50.77	Căramizi 5.73	Căramizi 10.88
	Pietre 52.30	Covor 47.62	Pietre 52.01	Catifea 5.79	Căramizi 10.89
Pietre.10	Covor 44.20	Blană 41.63	Pietre 45.24	Catifea 5.54	Gresie 10.94
	Blană 44.55	Covor 41.90	Blană 45.44	Scoarță 5.63	Catifea 10.95
	Blană 44.64	Blană 41.93	Blană 45.95	Gresie 5.68	Blană 10.96
Perete.10	Apă 50.40	Apă 44.92	Apă 48.47	Catifea 7.35	Pled 10.48
	Apă 50.62	Apă 45.02	Apă 49.83	Catifea 7.42	Perete 10.48
	Apă 50.63	Apă 45.38	Apă 49.91	Covor 7.42	Pled 10.48
Căramizi.10	Căramizi 50.34	Căramizi 49.19	Blană 55.19	Catifea 5.62	Perete 10.71
	Covor 53.57	Blană 49.82	Pietre 56.37	Catifea 5.73	Catifea 10.74
	Blană 53.65	Perete 50.08	Pietre 56.74	Catifea 5.82	Lemn 10.82
Sticlă.10	Covor 67.67	Covor 60.66	Covor 73.30	Catifea 6.58	Sticlă 10.52
	Pietre 68.98	Covor 61.13	Covor 73.83	Covor 6.64	Blană 10.63
	Covor 69.40	Covor 61.15	Covor 73.95	Covor 6.67	Sticlă 10.64
Covor.10	Gresie 60.18	Gresie 53.46	Gresie 58.53	Catifea 5.55	Blană 10.70
	Gresie 60.42	Gresie 53.73	Pietre 58.97	Covor 5.56	Scoarță 10.74
	Covor 60.45	Gresie 53.75	Căramizi 59.74	Pietre 5.65	Blană 10.75
Blană.10	Apă 39.20	Apă 35.80	Blană 43.59	Blană 5.64	Blană 9.89
	Blană 39.41	Blană 36.09	Blană 43.60	Catifea 5.96	Sticlă 10.38
	Apă 39.44	Apă 36.29	Blană 44.36	Căramizi 6.00	Blană 10.41
Catifea.10	Catifea 59.57	Catifea 48.11	Catifea 50.63	Catifea 4.86	Pled 10.64
	Catifea 61.63	Catifea 51.02	Catifea 54.88	Catifea 5.06	Lemn 10.69
	Catifea 65.02	Catifea 55.31	Catifea 58.43	Catifea 5.17	Pled 10.71
Pled.10	Perete 68.36	Pled 58.80	Pled 64.41	Pled 5.32	Pled 8.89
	Blană 68.66	Perete 60.94	Perete 65.39	Catifea 5.71	Pled 9.63
	Blană 68.87	Blană 61.01	Pled 65.82	Pled 5.77	Pled 10.05

Tabel II. Distanțele d_1, d_2, d_∞ cosinus și corelația Pearson aplicate coeficienților $dtcwt$ pentru prima textură din fiecare clasă rotită cu 15° comparată cu baza de date completă

	Distanțe				
	d_1	d_2	d_∞	cos_{DT}	r_{DT}
Scoarță.15	Scoarță 62.64	Scoarță 56.07	Scoarță 59.25	Catifea 5.61	Lemn 10.62
	Pietre 63.21	Scoarță 56.08	Gresie 60.27	Catifea 5.67	Pled 10.73
	Scoarță 63.46	Scoarță 56.14	Căramizi 60.28	Covor 5.67	Catifea 10.77
Lemn.15	Lemn 43.18	Lemn 34.36	Lemn 35.81	Catifea 5.74	Lemn 10.09
	Lemn 43.62	Lemn 34.47	Lemn 36.15	Catifea 5.85	Pled 10.12
	Apă 44.70	Lemn 34.74	Lemn 36.28	Catifea 5.98	Lemn 10.12
Apă.15	Apă 26.80	Apă 24.68	Apă 29.07	Catifea 5.80	Pled 10.29
	Apă 26.92	Apă 24.82	Apă 29.17	Apă 5.93	Sticlă 10.39
	Apă 26.94	Apă 25.21	Apă 29.26	Covor 6.00	Pled 10.44
Gresie.15	Pietre 51.14	Covor 47.51	Covor 50.98	Catifea 5.80	Sticlă 10.65
	Covor 52.14	Covor 47.79	Covor 52.17	Căramizi 5.88	Pled 10.78
	Pietre 52.23	Pietre 47.85	Pietre 52.31	Catifea 5.91	Pled 10.89
Pietre.15	Blană 44.43	Blană 41.77	Pietre 45.40	Catifea 5.49	Blană 10.86
	Pietre 45.07	Blană 41.83	Covor 45.42	Catifea 5.67	Covor 10.88
	Blană 45.09	Blană 41.93	Pietre 45.56	Scoarță 5.68	Pled 10.89
Perete.15	Apă 46.80	Apă 43.80	Apă 47.07	Catifea 7.28	Pled 10.43
	Apă 46.81	Apă 44.47	Apă 47.58	Catifea 7.29	Pled 10.46
	Apă 47.09	Apă 44.51	Apă 48.30	Gresie 7.31	Perete 10.58
Căramizi.15	Căramizi 51.56	Căramizi 48.84	Blană 54.46	Catifea 5.51	Lemn 10.80
	Pietre 53.41	Blană 49.49	Pietre 55.60	Catifea 5.68	Catifea 10.82
	Căramizi 53.74	Perete 50.15	Căramizi 55.70	Covor 5.74	Apă 10.84
Sticlă.15	Covor 68.83	Covor 61.31	Covor 73.03	Covor 6.62	Sticlă 10.48
	Covor 69.71	Covor 61.63	Covor 73.40	Catifea 6.62	Sticlă 10.57
	Covor 69.96	Covor 61.68	Pietre 73.76	Catifea 6.66	Catifea 10.60
Covor.15	Gresie 59.72	Gresie 53.56	Pietre 57.85	Catifea 5.52	Blană 10.72
	Gresie 60.19	Gresie 53.76	Gresie 58.75	Catifea 5.55	Blană 10.73
	Pietre 60.53	Gresie 53.92	Căramizi 59.14	Covor 5.55	Pled 10.78
Blană.15	Apă 39.15	Apă 35.15	Apă 42.67	Catifea 5.78	Blană 10.04
	Apă 39.37	Apă 35.40	Apă 43.49	Blană 5.81	Blană 10.40
	Apă 39.51	Apă 35.99	Apă 43.76	Catifea 5.93	Sticlă 10.46
Catifea.15	Catifea 57.19	Catifea 47.16	Catifea 50.44	Catifea 4.96	Apă 10.67
	Catifea 60.29	Catifea 51.71	Catifea 54.84	Catifea 5.00	Lemn 10.69
	Catifea 64.88	Catifea 55.10	Catifea 58.61	Catifea 5.07	Pled 10.71
Pled.15	Pled 66.73	Pled 58.25	Pled 62.03	Pled 5.46	Pled 8.97
	Căramizi 68.38	Pled 59.47	Pled 62.40	Catifea 5.89	Perete 10.03
	Blană 68.59	Blană 60.17	Blană 64.87	Catifea 5.90	Pled 10.08

De la percepția vizuală către detecția automată a texturii

Tabel III. Distanțele d_1, d_2, d_∞ cosinus și corelația Pearson aplicate coeficienților $dtcwt$ pentru prima textură din fiecare clasă rotită cu 10° comparată cu baza de date redusă

	Distanțe				
	d_1	d_2	d_∞	cos_{DT}	r_{DT}
Scoarță.10	Scoarță 63.39	Scoarță 54.64	Scoarță 58.79	Scoarță 5.72	Scoarță 10.73
	Scoarță 66.19	Căramizi 55.77	Căramizi 60.15	Catifea 5.74	Pled 10.83
	Căramizi 66.54	Scoarță 56.33	Scoarță 61.18	Catifea 5.76	Catifea 10.85
Lemn.10	Lemn 44.60	Lemn 33.22	Lemn 33.92	Catifea 5.61	Lemn 10.14
	Lemn 45.75	Lemn 33.86	Lemn 34.09	Catifea 5.81	Pled 10.15
	Lemn 46.37	Lemn 34.50	Lemn 35.67	Catifea 6.01	Catifea 10.34
Apă.10	Apă 25.76	Apă 24.37	Apă 28.84	Catifea 5.93	Apă 10.27
	Apă 26.36	Apă 24.42	Apă 28.85	Apă 6.00	Pled 10.57
	Apă 26.66	Apă 25.01	Apă 28.88	Catifea 6.06	Perete 10.58
Gresie.10	Covor 51.13	Covor 47.16	Covor 50.70	Catifea 5.72	Căramizi 10.88
	Pietre 52.30	Pietre 47.60	Pietre 52.01	Scoarță 5.89	Scoarță 10.94
	Pietre 53.43	Covor 47.62	Covor 52.90	Covor 5.91	Perete 10.98
Pietre.10	Pietre 44.99	Pietre 42.17	Pietre 45.24	Catifea 5.54	Covor 10.97
	Pietre 45.04	Covor 42.49	Blană 46.43	Scoarță 5.63	Catifea 11.01
	Covor 46.40	Blană 42.67	Covor 46.46	Catifea 5.74	Sticlă 11.02
Perete.10	Apă 51.34	Apă 45.41	Apă 49.91	Covor 7.42	Perete 10.48
	Apă 51.36	Apă 45.50	Apă 50.17	Catifea 7.43	Pled 10.62
	Apă 51.58	Apă 46.01	Apă 51.02	Catifea 7.49	Pled 10.78
Căramizi.10	Pietre 53.78	Pietre 50.92	Pietre 56.37	Catifea 5.62	Perete 10.71
	Pietre 54.06	Perete 51.20	Pietre 56.74	Catifea 5.82	Lemn 10.82
	Căramizi 54.63	Pietre 51.68	Căramizi 56.81	Covor 5.93	Catifea 10.95
Sticlă.10	Pietre 70.60	Pietre 61.87	Pietre 74.93	Covor 6.67	Sticlă 10.52
	Pietre 71.87	Pietre 62.46	Gresie 74.94	Catifea 6.72	Sticlă 10.64
	Covor 72.95	Covor 63.20	Covor 75.04	Apă 6.77	Lemn 10.85
Covor.10	Covor 60.45	Covor 54.90	Căramizi 59.74	Catifea 5.55	Scoarță 10.74
	Gresie 61.32	Căramizi 55.07	Covor 60.93	Covor 5.66	Perete 10.82
	Pietre 61.87	Căramizi 55.49	Căramizi 62.57	Catifea 5.70	Pled 10.89
Blană.10	Blană 39.41	Blană 36.09	Blană 43.59	Blană 5.64	Blană 9.89
	Blană 39.78	Apă 36.80	Blană 43.60	Catifea 5.96	Pled 10.51
	Blană 39.88	Blană 36.88	Blană 44.36	Catifea 6.00	Blană 10.56
Catifea.10	Catifea 59.57	Catifea 48.11	Catifea 50.63	Catifea 4.86	Pled 10.71
	Catifea 61.63	Catifea 51.02	Catifea 54.88	Catifea 5.06	Catifea 10.73
	Catifea 67.06	Catifea 55.93	Catifea 60.09	Catifea 5.19	Lemn 10.78
Pled.10	Perete 68.36	Pled 58.80	Pled 64.41	Pled 5.32	Pled 8.89
	Căramizi 69.78	Perete 60.94	Perete 65.39	Pled 5.77	Pled 9.63
	Pled 70.21	Pled 61.11	Pled 65.82	Catifea 5.79	Perete 10.09

Tabel IV. Distanțele d_1, d_2, d_∞ cosinus și corelația Pearson aplicate coeficienților dt_{cwt} pentru prima textură din fiecare clasă rotită cu 15° comparată cu baza de date redusă

	Distanțe				
	d_1	d_2	d_∞	\cos_{DT}	r_{DT}
Scoarță.15	Scoarță 63.46	Scoarță 56.08	Scoarță 59.25	Catifea 5.61	Pled 10.73
	Scoarță 63.74	Scoarță 56.14	Scoarță 61.05	Scoarță 5.68	Catifea 10.77
	Căramizi 64.66	Căramizi 56.58	Covor 61.37	Catifea 5.71	Lemn 10.82
Lemn.15	Lemn 46.00	Lemn 34.47	Lemn 36.15	Catifea 5.74	Lemn 10.32
	Lemn 47.95	Lemn 35.38	Lemn 37.19	Catifea 5.98	Catifea 10.40
	Lemn 48.09	Lemn 35.81	Lemn 37.36	Catifea 6.09	Catifea 10.44
Apă.15	Apă 26.80	Apă 24.68	Apă 29.07	Catifea 5.80	Apă 10.44
	Apă 26.94	Apă 24.82	Apă 29.37	Covor 6.00	Pled 10.48
	Apă 27.03	Apă 25.42	Apă 29.77	Scoarță 6.04	Pled 10.53
Gresie.15	Pietre 51.14	Covor 47.51	Covor 50.98	Catifea 5.80	Pietre 10.95
	Pietre 52.30	Covor 47.79	Covor 52.17	Scoarță 5.95	Căramizi 11.00
	Pietre 52.62	Pietre 47.85	Covor 52.43	Covor 6.01	Blană 11.01
Pietre.15	Pietre 45.07	Pietre 42.04	Pietre 45.40	Catifea 5.49	Blană 10.86
	Pietre 46.14	Blană 42.26	Covor 45.42	Scoarță 5.68	Pled 10.97
	Blană 46.95	Pietre 42.27	Pietre 45.56	Catifea 5.75	Catifea 11.00
Perete.15	Apă 47.09	Apă 44.87	Apă 48.96	Catifea 7.45	Pled 10.43
	Apă 48.41	Apă 44.88	Apă 49.28	Covor 7.52	Perete 10.58
	Apă 48.42	Apă 45.27	Apă 49.30	Covor 7.57	Perete 10.87
Căramizi.15	Pietre 53.41	Pietre 50.69	Gresie 57.60	Catifea 5.51	Apă 10.84
	Căramizi 53.74	Pietre 50.84	Pietre 57.83	Catifea 5.75	Pled 10.85
	Pietre 54.07	Perete 51.36	Pietre 57.89	Catifea 5.91	Lemn 10.88
Sticlă.15	Pietre 70.34	Pietre 62.66	Gresie 75.19	Catifea 6.66	Sticlă 10.48
	Pietre 71.27	Pietre 63.14	Covor 76.31	Catifea 6.73	Sticlă 10.57
	Gresie 72.66	Covor 63.97	Pietre 76.31	Catifea 6.85	Lemn 10.79
Covor.15	Gresie 62.68	Căramizi 55.03	Căramizi 60.61	Catifea 5.52	Scoarță 10.89
	Pietre 62.72	Covor 55.10	Gresie 60.83	Catifea 5.55	Apă 10.92
	Căramizi 62.86	Căramizi 55.12	Căramizi 60.86	Catifea 5.59	Catifea 10.93
Blană.15	Apă 39.37	Apă 35.99	Blană 43.99	Catifea 5.78	Blană 10.04
	Apă 40.00	Apă 36.03	Apă 44.26	Blană 5.81	Blană 10.40
	Apă 40.07	Apă 36.31	Blană 44.62	Catifea 5.94	Pled 10.51
Catifea.15	Catifea 57.19	Catifea 47.16	Catifea 50.44	Catifea 4.96	Pled 10.71
	Catifea 60.29	Catifea 51.71	Catifea 54.84	Catifea 5.00	Pled 10.71
	Catifea 66.35	Catifea 56.71	Catifea 60.81	Catifea 5.07	Blană 10.77
Pled.15	Pled 66.73	Pled 58.25	Pled 62.03	Pled 5.46	Pled 8.97
	Căramizi 68.38	Pled 59.47	Pled 62.40	Catifea 5.89	Perete 10.03
	Pietre 68.76	Căramizi 60.91	Perete 66.73	Covor 6.07	Pled 10.08

De la percepția vizuală către detecția automată a texturii

Tabel V. Distanțele d_1, d_2, d_∞ cosinus și corelația Pearson aplicate coeficienților $dtcwt$ pentru prima textură din fiecare clasă rotită cu 10° comparată cu baza de date cu o singură imagine în fiecare clasă

	Distanțe				
	d_1	d_2	d_∞	\cos_{DT}	r_{DT}
Scoarță.10	Scoarță 67.22	Scoarță 54.64	Scoarță 58.79	Scoarță 5.72	Scoarță 10.73
	Gresie 70.97	Pietre 59.20	Gresie 64.97	Catifea 5.76	Pietre 10.93
	Pietre 71.23	Gresie 60.46	Pietre 65.29	Gresie 5.97	Lemn 10.99
Lemn.10	Lemn 48.75	Lemn 34.50	Lemn 35.67	Catifea 6.01	Lemn 10.41
	Pietre 63.57	Catifea 49.92	Catifea 52.20	Lemn 6.31	Catifea 10.88
	Catifea 64.55	Apă 51.30	Pietre 54.72	Pietre 6.33	Pled 10.97
Apă.10	Apă 25.76	Apă 24.37	Apă 28.84	Apă 6.00	Apă 10.27
	Blană 42.84	Blană 38.72	Pietre 48.88	Catifea 6.13	Pled 10.68
	Pietre 47.76	Pietre 42.71	Blană 49.31	Gresie 6.33	Blană 10.79
Gresie.10	Pietre 52.30	Pietre 47.60	Pietre 52.01	Scoarță 6.05	Pled 11.02
	Gresie 53.63	Gresie 49.56	Gresie 54.64	Pietre 6.09	Căramizi 1.11
	Căramizi 7.14	Blană 50.99	Blană 58.18	Covor 6.11	Perete 11.12
Pietre.10	Pietre 45.04	Pietre 42.85	Pietre 47.18	Catifea 5.88	Gresie 11.08
	Blană 47.65	Blană 44.47	Blană 49.53	Scoarță 6.10	Lemn 11.10
	Apă 48.18	Apă 45.30	Gresie 50.18	Căramizi 6.19	Blană 11.10
Perete.10	Apă 53.40	Apă 46.14	Apă 51.99	Catifea 7.49	Lemn 10.80
	Blană 56.68	Perete 50.63	Perete 58.96	Pietre 7.61	Perete 10.94
	Pietre 56.96	Blană 50.95	Pietre 59.79	Covor 7.75	Pled 11.11
Căramizi.10	Pietre 53.78	Pietre 50.92	Pietre 57.50	Catifea 5.82	Apă 11.03
	Căramizi 5.28	Gresie 52.70	Gresie 58.53	Gresie 6.10	Lemn 11.04
	Gresie 55.45	Căramizi 3.23	Blană 61.73	Pietre 6.16	Pled 11.12
Sticlă.10	Pietre 71.87	Pietre 62.46	Pietre 75.58	Căramizi 6.87	Sticlă 10.52
	Gresie 73.83	Sticlă 64.93	Gresie 78.70	Catifea 6.96	Apă 10.94
	Blană 75.38	Gresie 64.98	Sticlă 79.84	Pietre 6.99	Blană 11.14
Covor.10	Pietre 65.67	Pietre 57.88	Scoarță 64.36	Catifea 5.70	Catifea 11.02
	Gresie 67.12	Gresie 59.51	Gresie 65.58	Covor 5.86	Pled 11.10
	Covor 67.97	Covor 59.67	Pietre 65.63	Pietre 5.95	Scoarță 11.12
Blană.10	Blană 39.78	Blană 36.09	Blană 43.59	Blană 5.64	Blană 9.89
	Apă 41.76	Apă 37.97	Apă 48.29	Catifea 6.06	Apă 10.75
	Pietre 50.05	Pietre 45.73	Pietre 52.52	Covor 6.19	Pietre 11.00
Catifea.10	Catifea 67.06	Catifea 55.93	Catifea 60.09	Catifea 5.06	Catifea 10.82
	Pietre 76.46	Lemn 64.07	Lemn 70.77	Covor 5.46	Pled 10.95
	Gresie 76.96	Scoarță 66.39	Scoarță 74.38	Căramizi 5.63	Covor 11.02
Pled.10	Blană 70.89	Pietre 62.75	Pietre 68.85	Pled 5.77	Pled 9.63
	Gresie 71.74	Blană 63.32	Pled 69.35	Scoarță 6.18	Catifea 10.82
	Pled 72.56	Apă 64.17	Gresie 71.25	Catifea 6.19	Apă 10.88

Tabel XII. Distanțele d_1, d_2, d_∞ cosinus și corelația Pearson aplicate coeficienților dt_{cwt} pentru imagini diverse comparate cu albumul Brodatz

	Distanțe			cos_{DT}		r_{DT}
	d_1	d_2	d_∞			
Scoarță1	Paie 88.24	Paie 75.14	Paie 79.10	Apă 5.82	Căramizi 10.98	
	Spumă 90.70	Paie 75.81	Piele 79.62	Lână 5.86	Nisip 11.00	
	Spumă 90.92	Piele 76.21	Paie 80.92	Nisip 5.86	Paie 11.01	
Scoarță2	Scoarță 88.84	Scoarță 76.31	Scoarță 79.69	Lână 5.63	Lână 11.02	
	Scoarță 89.00	Paie 76.60	Paie 81.07	Țesătură 5.64	Lână 11.03	
	Piele 90.86	Paie 77.02	Spumă 82.10	Pieleporc 5.72	Apă 11.05	
Scoarță3	Lână 67.40	Lână 62.08	Scoarță 69.46	Rafie 5.56	Căramizi 10.97	
	Pieleporc 69.25	Lână 62.36	Lână 70.61	Lână 5.60	Rafie 11.05	
	Lână 70.00	Lână 62.72	Pieleporc 71.08	Țesătură 5.60	Rafie 11.08	
Scoarță4	Pieleporc 41.21	Lână 37.19	Pieleporc 40.82	Iarbă 5.51	Apă 11.06	
	Lână 41.70	Apă 37.28	Lână 41.04	Țesătură 5.51	Paie 11.10	
	Rafie 41.81	Rafie 37.69	Rafie 42.41	Lână 5.52	Căramizi 11.11	

Tabel XII. Procentele de reușită a clasificării folosind cele cinci distanțe și cele patru baze de date

Distanțe	Baze de date			
	T completă	T redusă	T 1 imagine/clasă	Brodatz
d_1	50%	57,14%	46,43%	80%
d_2	57,14%	60,71%	50%	76,67%
d_∞	60,71%	60,71%	53,57%	76,67%
cos_{DT}	17,86%	25%	28,57%	73,33%
r_{DT}	32,14%	46,43%	53,57%	80%

Concluzii

Alături de estimarea formei și luarea în considerare a culorii, calculul texturii este deosebit de util în evaluarea similarității imaginilor, în adnotarea lor automată, într-o gamă foarte variată de aplicații.

Am comparat utilizarea bazei de date Brodatz cu 13 clase de texturi: apă, cărămizi, iarbă, lână, lemn, nisip, paie, piele, piele de porc, rafie, scoarță, spumă și țesătură cu utilizarea bazei de date Vistex, am restrâns gama variată de eșantioane, și am făcut comparații între eficacitatea metodelor utilizate.

Calculul texturii se dovedește a fi deosebit de laborios și consumator de resurse și timp, dar se dovedește în anumite cazuri a fi deosebit de util. Cercetarea inițiată continuă pentru a detecta procedura de evaluare a texturilor din imaginile sau înregistrările video, al căror număr crește veriginor, și pentru care sunt necesare procedurile automate de indexare, de adnotare.

Bibliografie

- Phil Brodatz, *Textures: A Photographic Album for Artists and Designers*, Dover Publications, New York, 1966, <http://www.ux.his.no/~tranden/brodatz.html>
 Arnaud Delorme, Ghislaine Richard, Michele Fabre-Thorpe, *Rapid Processing of Complex Natural Scenes: A Role for the Magnocellular Visual Pathways?*, Neurocomp., 26-27, 663-670, 1999.

De la percepția vizuală către detecția automată a texturii

- Michele Fabre-Thorpe, Arnaud Delorme et Ghislaine Richard, *Singes et Hommes face au monde visuel: la Catégorisation*, Primatologie, 2, 111-139, 1999.
- Michele Fabre-Thorpe, Arnaud Delorme, Catherine Marlot et Simon Thorpe, *A limit to the speed of processing in UltraRapid Visual Categorization of novel natural scenes*. Journal of Cognitive Neuroscience, 13, 171-180, 2000.
- Michele Fabre-Thorpe, Ghislaine Richard et Simon Thorpe, *Rapid categorization of natural images by rhesus monkeys*, Neuroreport, 9, 303-308. (1998).
- Denis Fize, Maxime Cauchois, and Michèle Fabre-Thorpe, *Humans and monkeys share visual representations*, Psychological and Cognitive Sciences, PNAS, Proceedings of the National Academy of Sciences, U S A, 2011 May 3, 108(18), pp. 7635-7640.
- Michal Haindl, Stanislav Mikes, *Texture Segmentation Benchmark*, Proc. of the 19th Int. Conf. on Pattern Recognition. IEEE Computer Society, 2008, pp. 1-4.
- Jean-Michel Hupé, Andrew C. James, Bertram R. Payne, Stephen G. Lomber, Pascal Girard, & Jean Bullier, *Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons*. Nature, 394, 784-787, 1998.
- Victor Lamme, Hans Supér, Henk Spekreijse, *Feedforward, horizontal, and feedback processing in the visual cortex*. Current Opinion on Neurobiology, 8, 529-535, 1998.
- Yan Liu, Rufin Vogels, and Guy A. Orban, *Convergence of Depth from Texture and Depth from Disparity in Macaque Inferior Temporal Cortex*, The Journal of Neuroscience, Behavioral/Systems/Cognitive, 24(15):3795-3800, April/14/2004.
- Svetlana Lazebnik, Cordelia Schmid, Jean Ponce - *A Sparse Texture Representation Using Local Affine Regions*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1265-1278, 2005.
- Marc J.-M. Macé, Arnaud Delorme, Ghislaine Richard, Michèle Fabre-Thorpe, *Spotting animals in natural scenes: efficiency of humans and monkeys at very low contrasts*, Animal Cognition May 2010, Volume 13, Issue 3, pp 405-418
- Bartlett W. Mel, SEEMORE: *Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition*, Neural Comp. 9, 777-804, 1997.
- Jerry D. Nguyenkim, Gregory C. DeAngelis, *Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons*. J. Neurosci. 23:7117-28, 2003.
- Ivan W. Selesnick, Richard G. Baraniuk, and Nick G. Kingsbury, *The Dual-Tree Complex Wavelet Transform*, IEEE Sign. Proc. Mag., vol. 22, no 6, pp 123-151, Nov. 2005.
- Simon Thorpe; Denis Fize; Catherine Marlot, Nature, „*Speed of processing in the human visual system*”, Nature, 1996 Jun 6; 381(6582): 520-522
- Russel L. De Valois & Karen K. De Valois, *Spatial Vision*, Oxford Univ. Press. N. Y., 1998.
- Niko Wilbert, *Hierarchical Slow Feature Analysis on Visual Stimuli and Top-Down Reconstruction*, <http://www.ini.rub.de/PEOPLE/wiskott/Reprints/Wilbert-2012-PhDThesis.pdf>, Doctoral thesis, Humboldt Univ., Berlin.
- Rui Xu, Donald C. Wunsch II, *Clustering*, IEEE Press/Wiley 2009.
- Wei Yu, J. Fritts, Fangting Sun, *A Hierarchical Image Segmentation Algorithm*, ICME, 2002.

Adrese web:

- Rubi Polikar, Wavelets Tutorials, <http://users.rowan.edu/~polikar/wavelets/wttutorial.html>
- Nick Kingsbury's personal website: <http://www-sigproc.eng.cam.ac.uk/~ngk/>
- Albumul cu texturi Brodatz - <http://sipi.usc.edu/database/>
- Baza de date de texturi VisTex - <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/>
- <http://taco.poly.edu/Wavelet Software/>
- http://www.sourcecodeonline.com/list?q=dual_tree_complex_wavelet_transform

Percepție, reprezentare și limitele posibilităților de cunoaștere

DAN GABRIEL SÎMBOTIN

ABSTRACT: The problem of perception is fundamental for a general epistemology. In the history of philosophy every analyses of the possibilities of knowledge had a foundation in the problem of perception of the reality. Even today this problem is still recent and unsolvable. The modern anatomy and psychology give a new vision of this problem, but is not enough. In relation with traditional epistemology and modern science, we try to explain the limits of perceptive realism, and how is possible to construct a unitary vision of reality.

KEYWORDS: *visual, perception, Gestalt, holism, imaginary*

Ca urmare a relației pe care o avem cu mediul, pentru care vom folosi pe viitor termenul de i-mediat¹, se construiește imaginea perceptuală și reprezentarea primară a lumii. Deoarece informația conștinetă, preluată din i-mediat este într-un procent considerabil (peste 80%) de natură vizuală, modalitatea în care funcționează văzul influențează în mod fundamnetal structura noastră cognitivă și modul în care integrăm ansamblul informațional. Problema pe care dorim să o dezbatem, pornind de la problema realismului, este în ce măsură i-mediatul este reflectat cu rigoare de către sistemul nostru perceptiv. Pentru aceasta vom urmări dacă imaginile perceptive au aceeași esență și structură cu imaginile mentale imparate, care este rolul subiectului în construcția imaginii perceptive, sau cu alte cuvinte imaginea perceptivă este una subiectivă (specifică fiecărui subiect) sau obiectivă (comună tuturor subiecților) și, în final, dacă nu putem considera chiar și imaginea perceptivă creație a propriului nostru eu. La acestea vom adăuga analiza problemei epistemice în care imaginea și în cazul cunoașterii științifice se păstrează elemente ale construcției vizuale.

1. Imaginea perceptivă recepție sau proiecție

În formarea imaginii perceptive, sistemul senzorial, specifice fiecărei specii joacă un rol fundamental în construcția imaginii i-mediatului. Pe lângă aceasta foarte

¹ Folosim acest termen „i-mediat” în locul celui de „realitate” deoarece dorim să eliminăm implicațiile ontologice pe care le are ca cel din urmă. I-mediatul este ceea ce intră în contact direct cu organele noastre de simț și pe bază căruia se construiește imaginea peceptivă și reprezentarea primară a lumii.

importantă este și personalitatea celui care preia mesajul. Atunci când ne referim la personalitate avem în vedere întreg complexul bio-psiho-social, incluzând atât aspectele cognitive cât și pe cele de ordin afectiv și organic. Personalitatea individului se proiectează asupra imaginii determinând variații ale acesteia. De aceea putem vorbi despre imagini ale lumii individualizate. Acești factori creează o imagine perceptivă diferită de la individ la individ, iar procesul perceptiv nu este un proces static, ci dimpotrivă unul dinamic și interactiv.

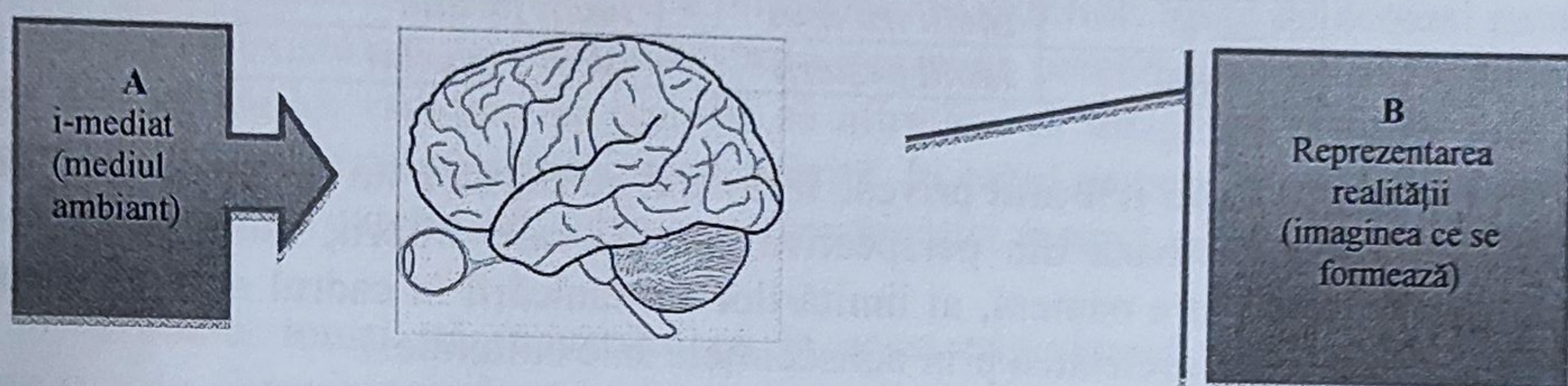


Fig. 1. Reprezentarea posibilităților de cunoaștere perceptivă

Pornind de la modelul cutiei negre (*black box*) folosit de behavioriști pentru reprezentarea procesului cognitiv considerăm că singurul element ferm cunoscut este unul subiectiv, și nu comportamental: propria imagine, pe care o considerăm o reprezentare a realității. Dar cât de diferite sunt imaginile individuale? Pot acestea să redea i-mediatul așa cum este el? Răspunsurile sunt diferite. Ne putem pronunța așa cum o face în sinteză Mielu Zlate. „Psihicul poate reproduce în sine, în interiorul său realitatea înconjurătoare așa cum este ea, dar și într-o formă modificată, așa încât realitatea din mintea omului să nu mai semene cu realitatea din afara minții lui.”² Această perspectivă moderată care încercă să împace realismul cu o formă de subiectivismul perceptiv, poate fi o soluție prin care nu sunt negate nici posibilitățile de eroare ale percepției dar nici capacitatea acesteia de a reda în contextul adecvat i-mediatul. Astfel B reflectă A iar distorsiunile care pot să apară sunt doar accidente.

Modalitatea în care este folosită descrierea psihicului scoate în evidență faptul că imaginea perceptivă nu este efectul unei construcții mentale, ci ea redă realitatea, în anumite momente cu distorsiuni care au drept cauză componenta subiectivă. Considerăm că această perspectivă apropiată de realismul clasic, care chiar dacă încearcă să rezolve problema erorilor de percepție, pe care nimeni nu le neagă, nu dă un răspuns în legătură cu proiecția imaginilor proprii asupra a ceea ce reprezintă imaginea perceptivă primară. Ea are o tradiție îndelungată, iar empirismul epocii moderne, prin viziune lui Francis Bacon, o sintetizează perfect „Se poate spune chiar, în aceasta privință, că mintea este o oglindă care primește și reflectează razele lucrurilor, nu pe un plan unic, ci pe o mulțime de fațete așezate diferit, așa încât nu există nimeni care, datorită educației, studiilor și naturii sale proprii, să nu fie sub înrâurirea unei puteri ademenitoare și ca pradă unui dușman obișnuit care îl înșeală și îi tulbură mintea printr-o mulțime de arătări deșarte.”³

² Mielu Zlate, *Introducere în psihologie*, Editura Polirom, Iași 2000, p. 222.

³ F. Bacon, *Cugetări și observări în Despre înțelepciunea antichilor*, Editura Științifică și Enciclopedică, București, 1976, p. 265.

„Metafora oglinzii” care distorsionează a realitatea a fost ulterior metamor-
fizată în cei patru idoli⁴ care surprind toate direcțiile cunoașterii omenești. Cei patru
idoli au fost divizați după următoarele criterii: cantitativ – omul ca particular
(individul) și omul ca universal (societate) și calitativ (cultura) – omul în stare
naturală și cel cultural. Făcând o sinteză a idolilor îi putem reprezenta astfel:

Om	Individ	Societate
Natural	<i>Idolii tribului</i>	<i>Idolii forului</i>
Cultural	<i>Idolii peșterii</i>	<i>Idolii teatrului</i>

Sintetic, **idolii tribului** privesc individul cu puterile lui de cunoaștere, cei **ai
peșterii** privesc persoana din perspectiva culturii sale proprii, **idolii forului** sunt
idolii legăturilor dintre oameni, ai limitărilor comunicării în cadrul societății, iar cei
ai teatrului privesc societatea prin consecințele sale culturale.

Acest exemplu arată că dacă preferăm varianta teoretică prin care **B** este o
reproducere mai mult sau mai puțin fidelă a i-mediat-ului, atunci putem identifica o
multitudine de elemente care intervin între în constituirea erorilor. Din punctul nostru
de vedere mai interesantă și cu un potențial explicativ mai ridicat este perspectiva
dezvoltată de Christensen și Klivington. Aceștia consideră că rolul creierului nu este
acela de a filma realitatea, ci mai degrabă „o reconstituie și într-o oarecare măsură o
și creează”⁵.

Imaginea pe care o avem despre lume din perspectiva noastră nu constituie
„realitatea” ca o redare fidelă a i-mediatului, ci o formă subiectivă de decodare a
acesteia subiectivă. Deoarece formarea imaginii generale începe de la ceea ce se
poate numi simț comun care nu reprezintă realitatea într-un mod „obiectiv” trebuie
identificată sursa/sursele imaginii perceptive. Dacă imaginea perceptivă nu reprezintă
realitatea, ci este o construcție, ea aparține nivelului imaginar, ea existând doar ca
imagine mentală și prin aceasta subiectivă. De aceea trebuie să considerăm că
imaginarul este o formă integrantă a „realității” și „realitatea” formă integrantă a
imaginarului.

Astfel, pentru susținerea acestei idei este necesar să identificăm dacă
imaginea senzorială reprezentarea are aceeași structură ca reprezentările mentale.
Imaginea formată prin senzații nu reprezintă un ansamblu holistic al spațiului încon-
jurător. Ea este limitată, iar orice explicație exhaustivă a acesteia este o construcție
subiectivă. Ne construim propria lume prin forme de specifice de operare a propriului
nostru creier: analiza, generalizarea, abstractizarea etc. Imaginile generale asupra
lunii sunt construcții proprii, dar este posibil ca realitatea să fie într-adevăr așa cum
o percepem? Să fie lumea construită pe baza unor mecanisme care redau mediul
asemenea unei oglinzi? Imaginea mentală este tot una cu acea parte din i-mediat pe
care o sesizăm. Sau este posibil să proiectăm propriul imaginar asupra realității
înconjurătoare și astfel să fim proprii constructori ai realității pe care o percepem?

Nu există diferențe majore între imaginile mentale ce se realizează ca urmare
a percepției și cele ce nu mai au ca suport obiectul. Pentru a ne susține punctul de

⁴ Fr. Bacon, *Noul Organon*, I, XLI-XLIV, Editura Academiei, București, 1957, pp. 42-57
⁵ Mielu Zlate, *Introducere în psihologie*, ed. cit, p. 222.

vedere vom descrie experimentul realizat de Perky în 1910.⁶ În fața unui ecran translucid marcat cu un punct în centru experimentatorul așează o persoană ce trebuie să-și focalizeze mental atenția în centru și să-și imagineze o roșie. Pe spatele ecranului este proiectat conturul unei tomate în culoare roșie la o intensitate inferioară celei perceptive. Treptat intensitatea crește până când devine perceptibilă. Subiectul supus experimentului nu realizează nici o diferență, el declarând în continuare că roșia este imaginată. Putem concluziona că nu există nici o diferență de substanță între imaginea proiectată mental și cea receptată din exterior. Fiind de aceeași natură este posibil să existe o interrelaționare între imaginea cerebrală și cea recepționată. Lipsa determinărilor radicale ne permite să afirmăm că imaginile mentale sunt de aceeași structură, sursa impulsului fiind diferită. În cazul percepției energia ce determină imaginea este exterioară, iar în cazul imaginației aceasta are drept sursă energia interioară.

Trebuie, totuși, să remarcăm că în acest experiment s-au folosit imagini foarte simple. Oare imaginile mentale nu sunt asemănătoare celor reale doar în cazul imaginilor simple? Dacă în cazul imaginilor complexe asemănarea nu mai este valabilă? Pentru a realiza analiza imaginilor complexe vom interpreta două experimente realizate ulterior celui al lui Perky. Primul este cel realizat de Shepard și Metzler⁷, 1976, în care pe un ecran erau prezentate diferite două figuri geometrice una lângă cealaltă. Acestea reprezentau aceeași imagine în unghiuri diferite. Era cronometrat timpul până când persoana asupra căreia era efectuat experimentul realiza diferențele dintre cele două figuri. S-a observat că perioada de timp este direct proporțională cu unghiul de rotație al imaginii. Comportamentul era de așa natură de parcă imaginile ar fi avut rigiditate fizică și o viteză de rotație măsurabilă. Acest experiment arată că la nivel mental avem tendința de realiza aceleași mișcări cu deprinderile pe care le avem cu privire la lumea exterioară. Pentru noi nu există diferențe foarte mari de structură între cele două tipuri de imagine.

Al doilea experiment realizat de Kosslyn⁸ în 1980 s-a concretizat cu niște concluzii asemănătoare celor prezentate anterior. Experimentul consta în realizarea unei expediții imaginare. Inițial subiectul este rugat să deseneze harta unei insule. Ulterior, în lipsa hărții acesta este să-și imagineze că se află pe plajă, iar experimentatorul îi cere să identifice anumite obiecte de pe insulă. Este măsurat timpul necesar identificării, iar acesta este direct proporțional cu distanța de la plajă la obiect de parcă persoana parcurgea în imaginație drumul până la obiect. Harta mentală conținea aceleași informații ca și cea reală iar identificarea se realiza de parcă obiectul era căutat pe o hartă reală. Cele două experimente ne-au dus la concluzia că imaginea mentală este asemănătoare celei reale, iar activitățile ce aparțin imaginației sunt de același tip cu cele ce aparțin gândirii sau identificării perceptive. Diferențele dintre imaginație, percepție și alte procese cognitive nu sunt de esență, ci doar dezvoltate pe un fundament teoretic diferit.

Din cele precizate mai sus se poate concluziona că imaginea perceptivă și cea imaginată au aceeași structură, diferențele dintre ele fiind de sursa acestora. Dar

⁶ Jean Pierre Changeaux, *L'Homme neuronal*, Fayard, Paris, 1983, pp. 177-178.

⁷ *Idem* p. 175.

⁸ *Ibidem* p. 175.

întrebarea ce apare acum este următoarea: joacă vreun rol omul în construcția imaginii perceptive, care constituie baza pentru imaginea despre lume pe care o numim realitate? Răspunsul pe care dorim să-l dăm este pozitiv și pentru aceasta în primul rând vom prezenta ceea ce J Rock susține în cartea sa *The logic of perception*⁹: percepția este inteligentă. De fapt suntem obligați să precizăm că nu percepția este inteligentă, ci modul în care se formează imaginea perceptivă este inteligent. Această modalitate de construcție a imaginii se datorează unei modalități complexe de relaționare între procesele psihice. De aceea atunci când discutăm despre psihicul uman trebuie să-l privim în mod holistic, neexistând rupturi între procesele psihice.

2. Holismul și formarea imaginii despre lume

Regula de bază după care funcționează percepția constă în realizarea unei imagini holistice asupra lumii pornind de la imagini individuale sau frânturi de imagini. În primul rând imaginea perceptivă generală se formează pornind de la imaginile individuale care se află la nivelul „ferestrei atenției” (*attention window*)¹⁰ prin combinarea acestora astfel încât să se obțină o imagine de ansamblu asupra realității. Această tendință de a integra într-un tot unitar imaginea este dublată de proiecția asupra imaginii percepute a imaginilor deja existente în memorie. Aceste două elemente sunt fundamentale pentru ceea ce am numit „inteligenta percepției”. Pentru a susține această idee trebuie urmărită modalitate în care se recunoaște obiectul în cazul imaginilor incomplete.

Experimentele realizate în această direcție au subliniat modalitatea în care se proiectează imaginea existentă în memorie asupra obiectului perceput. Sintetizând aceste experimente se pot enumera următoarele tipuri¹¹:

1. „Contururi degradate” realizate de Biederman și Binckle în 1985 reprezintă experimentele în care s-au eliminat părți din conturul unui obiect, încercându-se identificare acestuia. Lucrul acesta a fost posibil datorită imaginilor deja existente la nivel mental.
2. „Părți lipsă” reprezintă experimentele în care au fost mutate părți dintr-un anumit obiect și s-a urmărit posibilitățile de identificare ale obiectului. Astfel de experimente au fost realizate de Biederman 1987, Cave și Kosslyn 1993, Biederman și Cooper 1991, 1992.
3. „Imagini fărâmițate” sunt experimentele în care imaginea era fragmentată, trebuind reconstituită. Experimentele realizate de Cave și Kosslyn 1993 au arătat că există tendința de a asocia părțile aflate în apropiere, și nu cele ce se potrivesc din punct de vedere logic. Dar în urma procesului de verificare cu imaginea deja existentă, acestea au fost reorganizate și s-au realizat, în cele din urmă, imagini coerente.
4. „Întreruperi spațiale între părți” constituie experimentele (Cave și Kosslyn 1993) în care s-au realizat mutări între părțile componente ale unor obiecte. Obiectele

⁹ J. Rock, *The logic of perception*, M.I.T. Press, Cambridge, 1983.

¹⁰ Stephen M. Kosslyn, *Image and Brain*, Bradford, Cambridge Massachusetts, 1995, pp. 187-194.

¹¹ *Idem* p. 265.

erau reconstituite pe baza imaginii ce se considera că este cea inițială. Această imagine putea fi identificată și pe baza direcționări verbale.

Aceste experimente au determinat o anumită perspectivă asupra imaginii perceptive sintetizată în legile gestaltiste ale percepției:

- „a) *principiul proximității* – elementele aflate în proximitatea spațială sunt grupate într-o singură unitate perceptivă;
- b) *principiul similarității* – elementele similare sunt grupate în aceeași unitate perceptivă, care e contrapusă altora;
- c) *principiul buneicontinuări* – la intersecția a două contururi ele sunt percepute după continuarea cea mai simplă;
- d) *principiul închiderii* – conturul ocluzat al unei figuri este închis după configurația sa vizibilă.”¹²

Atât experiențele descrise mai sus cât și legile gestaltiste pornesc de la ideea enunțată deja că percepția are tendința de a uni imaginea, creând o structură holistică. Datorită acestei tendințe spre unificare, folosindu-ne de imaginile deja existente, subiectul devine un element important în a crea realitatea înconjurătoare. Imaginea exteriorului este o construcție între ce are la bază relația subiectiv – obiectiv.

Putem afirma că relația dintre eu și mediu este una interactivă, omul preluând din mediu impulsul energetic, construiește imaginea sa mentală care este subiectivă. Din acest punct de vedere putem considera în conformitate cu Ittelson că percepția este: „o tranzacție, un fel de interrelație sau schimb între organism și mediu, dar în care fiecare parte a situației intervine ca participant activ și își datorează existența tocmai acestei participări active.”¹³ Interacțiunea dintre mediu și subiect nu se manifestă în mod necesar prin modificarea mediului, dar imaginea percepută nu reprezintă preluarea identică a acestuia. Unitatea dintre universul gnoseologic și cel perceptiv este foarte strâns, astfel încât imaginea perceptivă este formată pe baza celor două, neputându-se realiza o ruptură radicală în interiorul omului. Imaginile umane sunt de același fel indiferent de sursa lor, iar imaginea despre lume este una și ea este parte integrantă a psihicului uman. Aici am ajuns la problema larg dezbătută, pe la mijlocul secolului al XX-lea, a diferenței dintre cunoașterea specifică simțului comun și cea științifică. Noi nu susținem că cele două de cunoaștere ar determina aceleași imagini, ci dimpotrivă construcția lor se realizează în mod diferit determinând imagini diferite.

3. Construcția imaginii despre lume

Totuși tendința din psihologia teoretică este de a împărți psihicul uman între componentele cognitive și cele perceptive, între aceste două elemente existând o strânsă legătură. Nivelul cognitiv influențează modul în care imaginea se formează fiind un fel de matrice a acesteia. Există un număr mare de experimente care au avut

¹² Mircea Miclea, *Psihologie cognitivă*, Editura Polirom, Iași, 1999, p.82.

¹³ Ittelson, *Visual space perception*, apud Mielu Zlate, *Psihologia mecanismelor cognitive*, Editura Polirom, Iași, 1999, p. 129.

drept scop identificarea relației dintre imagine și cuvânt. Dintre acestea credem că cel mai reprezentativ este cel al psihologilor Carmichael, Hogan și Walter, 1932¹⁴. Aceștia au prezentat la două la grupuri de subiecți o serie de imagini ambigue. În cadrul primului grup imaginile au fost însoțite de diverse denumiri. Tendința celor din grupul supus experimentului a fost de a recunoaște imaginea care le-a fost indusă de stimulul verbal. Astfel datorită asociației ce se realiza între imagine și cuvânt acestea erau deformatе în direcția cuvântului asociat. De ce se realizează acest lucru? Explicația noastră este următoarea: cuvântul proiectează la nivelul mental al individului o imagine. Această imagine este imaginea necesară pentru a dirija percepția așa cum am observat în experiențele anterioare. Atunci când imaginea generală este considerată ca fiind cunoscută, atunci imaginea perceptivă se identifică cu aceasta. În cadrul acestui experiment cuvântul joacă doar un rol de intermediar între imaginea mentală și cea perceptivă.

În cazul imaginilor duble putem percepe obiecte diferite uitându-ne la aceeași imagine pe baza unor criterii subiective. Aceste imagini duble subliniază faptul că subiectul și proiectează propriile stări, propria structură cognitivă asupra imaginii pe care o percepe. Percepția este un fenomen subiectiv, nu numai prin faptul că sistemul receptor este propriu fiecăruia ci, și prin faptul că imaginea ce se formează la nivel mental este o sinteză între structura cognitivă afectivă proprie și elementele exterioare. Datorită acestei relații ce există între cognitiv și perceptiv a apărut problema sistemului de imagini complexe care sunt specifice fiecărei persoane. Urmărind sistemul cognitiv și modalitatea de activare a acestuia, am ajuns la concluzia unei conștientizări parțiale ale proceselor psihice. „Fereastra atenției” este mult prea îngustă pentru a surprinde și susține activitatea cognitivă specifică fiecăruia. Deprinderile sunt cele care operează în majoritatea acțiunilor omenești, conștientizarea acestor acțiuni fiind redusă sau chiar lipsind. Același fenomen este întâlnit și în cazul percepției, deoarece imaginea percepută se realizează pe baza cunoștințelor deja existente, prin proiecția imaginii cerebrale asupra structurii mai mult sau mai puțin complete a imaginii percepute. Mai mult înțelegerea și integrarea imaginii se face tot pe baza unor construcții deja existente. O imaginea ce nu se integrează în acest sistem este percepută ca imagine dilematică, și neintegrată în sistemul imaginar propriu.

Am argumentat până în acest moment faptul că imaginea perceptivă este formată prin intervenția structurii cognitive specifice fiecăruia. Dar de ce apare această implicare și de ce imaginea cognitivă nu este copia fidelă a realității? În primul rând pentru că există mai multe etape în trecerea de la energia exterioară la imaginea noastră mentală.

1. Prima dintre etape este cea care a fost numită *image activation*¹⁵ (activarea imaginii) și constă în două procese complementare. În primul rând putem vorbi despre o reprezentare de suprafață proces ce determină formarea *imaginii primare*¹⁶. Aceasta reprezintă o schiță a imaginii ce urmează a fi percepută, în această fază fiind extrase contururile. În același timp are loc și un proces de

¹⁴ Mielu Zlate, *Psihologia mecanismelor cognitive*, Editura Polirom, Iași, 1999, p. 128.

¹⁵ Stephen M. Kosslyn, *Image and Brain*, Bradford, Cambridge Massachusetts, 1995, p. 146.

¹⁶ Mircea Miclea, *Psihologie cognitivă*, Ed. Polirom, Iași, 1999, pp. 66-75 și Jean Pierre Changeaux, *L'Homme neuronal*, Fayard, Paris, 1983, pp. 187.

profunzime în care sunt activate imaginile comparative cu imaginea primară. Pe baza acestor procese se formează *imaginea secundară*, cea prin care conturul devine imagine de ansamblu.

2. A doua etapă este cea denumită *image inspection*¹⁷ (inspectarea imaginii) și este reprezentată de procesul prin care sunt surprinse detaliile imaginii. În această etapă intervin deja procesele cognitive pentru că o parte a detaliilor pot fi și construite.
3. A treia etapă constă în *image maintenance*¹⁸ (menținerea imaginii) și reprezintă procesul în care imaginea perceptivă este stabilă și recunoscută. Procesul este complex și se realizează pe baza relației dintre cognitiv și perceptiv.
4. Ultima etapă este *image transformation*¹⁹ (transformarea imaginii) și constă în transformarea imaginii pentru a putea fi integrată în complexul cognitiv propriu. Această etapă este cea care determină construcția imaginarului ca imagine de ansamblu. Fiecare dintre imaginile ce se integrează trebuie să fie compatibile cu imaginarul persoanei, altfel ele sunt marginalizate și se constituie în imaginarul marginal sau pur și simplu sunt eliminate din structura de imagine.

Aceste etape sunt necesare pentru a integra imaginile noi între celelalte imagini și a alcătui imaginarul unei persoane. Fiecare dintre etape joacă rolul său în constituirea imaginarului perceptiv.

Dar majoritatea experimentelor descrise au avut drept punct de plecare imaginile simple statice. Există aceleași procese și la nivelul imaginilor complexe? Dar ce reprezintă mișcarea și cum se formează ea la nivel mental?

Pentru a răspunde la aceste întrebări un element care trebuie urmărit este reprezentat de percepția succesiunii. Și în acest caz putem urmări intervenția subiectivului asupra „realității obiective”. Tot la nivelul percepției mișcării și a succesiunii trebuie urmărită și problema percepției temporale. Problema care apare aici este dacă percepția succesiunii reprezintă o formă a „realului” sau o construcție mentală. Una dintre caracteristicile principale ale mișcării este datorată inerției imaginii. Astfel imaginile statice ce se succed cu o viteză mai mare de 1/25 secunde dau senzația mișcării. Această inerție a sistemului senzorial poate constitui sursa unor variații perceptuale astfel încât imagini statice pot să creeze o mișcare aparentă.

Astfel trebuie urmărită mișcarea din mai multe puncte de vedere. Mielu Zlate realizează o clasificare a percepției mișcării identificând cinci tipuri de percepție: *reală* (când obiectul se mișcă dintr-un loc în altul); *aparentă* (când obiectul pare că se mișcă chiar dacă este nemișcat); *indusă* (când obiectul este inclus sau lângă un obiect în mișcare și pare a se mișca); *autocinetică* (când privim un punct luminos într-o cameră întunecată fără alt punct de reper punctul respectiv pare în mișcare); *consecutivă* (după ce urmărim un obiect în mișcare, dacă urmărim altul și acesta va părea în mișcare)²⁰. Aceste tipuri de percepție a mișcării arată tendința psihicului de a crea continuitate, legături între elementele componente ale percepției. Această ten-

¹⁷ Stephen M. Kosslyn, *op. cit.*, p. 149.

¹⁸ *Idem*, p. 150.

¹⁹ *Ibidem*, p. 151.

²⁰ Mielu Zlate, *Psihologia mecanismelor cognitive*, Editura Polirom, Iași, 1999, p. 157.

dință de continuitate nu se bazează doar pe inerțialitatea sistemului senzorial ci și pe arată intervenția cognitivului în percepția mișcării.

Timpul subiectiv este tot o formă de percepție a mișcării. Acesta este reprezentat de viteza aparentă de mișcare, care se constituie sub forma unui raport dintre viteza de acțiune proprie la un moment dat și viteza de acțiune medie. În absența percepției mișcării nu ar exista timp, pentru că acesta este măsură a mișcării. Tot raportată la mișcare este și cauzalitatea. Fără a exista mișcare nu putem identifica legături între obiecte, evenimente, imagini. O percepție flash ar constitui sursa unei lumi fără timp și relații cauzale. Aceasta ar fi o lume ruptă un perpetuu moment prezent. De asemenea fără ideea de cauzalitate și temporalitate nu există o psihologie a mișcării. Relația dintre cele percepția mișcării și timp este una de influență și susținere reciprocă.

Totuși vorbind despre constituirea imaginarului perceptiv și despre modalitate de interrelaționare dintre imaginea cognitivă deja existentă și imaginea percepută se conturează problema modalităților de constituire a imaginarului existent. Ce informații sau structuri informaționale deținem la naștere, ce învățăm și ce se află deja în noi reprezintă întrebările la care încercăm să răspundem în continuare. Nu vom încerca și nici nu putem să rezolvăm acum problema surselor cunoașterii. Totuși vom încerca să prezentăm modalitatea în care se formează imaginea despre lume a fiecărei persoane. Problema aceasta poate fi analizată pornind de la psihologia genetică a lui J. Piaget. Raportul dintre percepție și universul cognitiv este unul constructiv, ce se dezvoltă în timp. El pornește de la o concepție empiristă pură eliminând ideea că la naștere ar exista orice formă de cunoaștere. El consideră că nici măcar intuițiile apriori de spațiu și timp²¹ nu sunt specifice omului. Jean Piaget consideră spațiul și timpul drept construcții perceptuale ce se realizează în copilărie. „Abordând mai întâi structurile spațio-temporale, constatăm că la început nu există nici un spațiu unic, nici o ordine temporală, care să înglobeze obiectele și evenimentele așa cum recipientul înglobează conținutul său. E dat doar un ansamblu de spații eterogene, toate centrate pe corpul propriu al copilului: spațiul bucal, spațiul tactil, vizual, auditiv, postural, precum și unele impresii temporale (așteptarea etc.) dar fără coordonări obiective.”²² Acest spațiu și timp care se dezvoltă în interior vor fi exteriorizate odată cu alte elemente specifice.

Astfel realitatea obiectivă este o construcție subiectivă ce va fi obiectivată treptat, odată cu dezvoltarea copilului. „Studiul inteligenței senzorio-motorii sau practice din cursul primilor doi ani de dezvoltare ne-a arătat cum copilul, începând prin a asimila direct mediul exterior activității sale, construiește apoi, pentru a prelungi această asimilare, un număr tot mai mare de scheme, în același timp mai mobile și mai apte de a se coordona între ele.”²³ Concepția genetică a lui Piaget constituie o modalitate prin care putem urmări modalitățile de dezvoltare specifice imaginarului. Constituirea „inteligenței” se face treptat, învățarea se realizează gradual și prin aceasta imaginea despre lume se construiește pas cu pas. Lumea în care trăim este o lume a subiectivității, imaginarul se construiește în interior prin

²¹ Immanuel Kant, *Critica Rațiunii pure*, Editura IRI, București, 1994, pp. 71-94.

²² Jean Piaget, Barbel Inhelder, *Psihologia copilului*, Editura Didactică și Pedagogică, București, p. 16.

²³ Jean Piaget, *Constituirea realului la copil*, Editura Didactică și Pedagogică, București 1976, p. 3.

relaționare cu mediul natural și social și ulterior o proiectăm în afara, ca formă obiectivă a realității. Realitatea constituită interior ca imagine mentală este proiectată ulterior ca obiect: „este limpede că recunoașterea nu conduce în nici un caz de la sine și fără complicații ulterioare la noțiunea de obiect.”²⁴ Constituirea mentală a obiectului se va realiza treaptă cu treaptă în funcție de vârstă mintală a copiilor. Dezvoltarea universului cognitiv implică și dezvoltarea universului perceptiv, pe baza schemelor pe care copilul va învăța să acționeze. Chiar dacă în concepția sa Piaget este un realist, nepunând nici o clipă la îndoială realitatea înconjurătoare, teoria și experimentele sale pot susține un subiectivism perceptiv total. Nu este obligatoriu ca mediul natural să fie în forma în care este perceput de noi, este suficient suportul mediului social care proiectează o realitate naturală convenită.

De aceea este posibil să considerăm percepția doar ca o construcție subiectiv-imaginativă în care fiecare dintre imaginile individuale ale noastre și ale celorlalți își aduc aportul. Făcând o sinteză a modului în care se formează imaginarului perceptiv ajungem la următoarele concluzii. Imaginea perceptivă este sinteza imaginilor vizuale, auditive, senzorio-motorii și kinestezică. Toate formele de percepție sunt diferite, ele apărând în funcție de distanță în momente diferite, dar sunt asamblate într-o singură imagine perceptivă. Astfel psihicul intervine în a realiza o imagine de ansamblu a realității prin intermediul nivelului cognitiv care participă în mod activ la formarea acesteia. Faptul că imaginarul nu trebuie redus la ideea de imagine vizuală poate fi susținută prin modalitățile de a percepe realitatea specifice orbilor. Aceștia au o imaginație de ordin imagistic, chiar dacă nu au perceput niciodată o imagine vizuală. Construcția lor s-a realizat într-un mod deformat, dar prin intermediul celorlalte simțuri au reușit să-și construiască o imagine despre lume proprie. Diferențele dintre imagini dintre un om normal și un orb nu sunt de esență ci de acuratețe, ceea ce arată că imaginea este una dintre modalitățile fundamentale de operare ale sistemului cognitiv. De fapt imaginea este considerată de către noi drept cărămida de bază a psihicului uman ea fiind modalitatea de a transmite, structura și crea cunoștințe.

Imaginarul se construiește pe baza imaginii perceptive. Aceasta se formează prin interrelaționarea subiect obiect ce are drept bază anumite cerințe de construcție. Acestea ar putea fi sintetizate astfel:

1. Imaginea perceptivă este influențată de imaginarul deja existent al fiecăruia, construindu-se pe baza relației subiect – obiect.
2. Construcția imaginarului fiecăruia se realizează într-un mod treptat, natural, generându-se un ansamblu vital integrării spațio-sociale. Orice modificare bruscă a imaginarului determină stări de disconfort personal putându-se ajunge până la traume.
3. Există tendința psihicului uman de a construi mai multă ordine, chiar acolo unde aceasta nu există. Aceasta determină o imagine unitară a cosmosului, iar acolo unde există inadvertențe acestea sunt adaptate ordinii personale sau imaginile sunt respinse.
4. Psihicul funcționează pe baza imaginii holistice, „interpretarea” lumii jucând un rol foarte important în „formarea” ei.

²⁴ *Idem*, p.10.

5. Percepția este un ansamblu realizat din imaginile vizuale, auditive și senzorio-motorii strâns legate între ele. Ele se finalizează prin intermediul unei imagini de ansamblu asupra i-mediatului.

Deoarece imaginarul constituie principala formă de manifestare a universului cognitiv este necesar să urmărim și relația ce există între personalitatea umană ca ansamblu și imaginarul individual.

În loc de concluzii: eroarea perceptivă ca fundament al reproducerii imaginii

Un rol important în dezvoltarea teoretică îl joacă reconstrucția și reproducerea imaginii. Astfel, imaginea spațială prin transferul de la bidimensional la tridimensional se fundamentează pe modalitatea în care percepem unghiul și imaginea holistică, respectând principiile gestatiste. Pictura ca reprezentare a realității se fundamentează ca modalitate de reconstituire a vizualului pornind de la reflexiile luminoase și modalitatea în care acestea redau culorile, lumina și umbrele. Vizualul este redat prin limintele pe care le are propriul nostru aparat receptor, și de aceea orice reproducere a imaginii trebuie speculeze aceste limite.

Bibliografie

- Bacon, Francis, *Despre înțelepciunea anticilor*, Editura Științifică și Enciclopedică, București, 1976.
Bacon, Francis, *Noul Organon*, Editura Academiei, București, 1957.
Changeaux, Jean Pierre, *L'Homme neuronal*, Fayard, Paris, 1983.
Rock, John, *The logic of perception*, M.I.T. Press, Cambridge, 1983.
Kosslyn, Stephen M., *Image and Brain*, Bradford, Cambridge Massachusetts, 1995.
Miclea, Mircea, *Psihologie cognitivă*, Editura Polirom, Iași, 1999.
Piaget, Jean, Inhelder, Barbel, *Psihologia copilului*, Editura Didactică și Pedagogică, București.
Piaget, Jean, *Constituirea realului la copil*, Editura Didactică și Pedagogică, București, 1976.
Zlate, Mielu, *Introducere în psihologie*, Editura Polirom, Iași, 2000.
Zlate, Mielu, *Psihologia mecanismelor cognitive*, Editura Polirom, Iași, 1999.

Index

- Algoritm
- de factorizare, 187, 191
 - de formare a clusterelor, 22, 23, 31, 203
 - Runge-Kutta, 21
- cameră afină, 184, 188, 189, 201
- clasificare, 12, 24, 63, 167, 169, 172, 261, 262, 270, 291
- contrast, 105, 106, 247, 252
- corelația Pearson, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282
- corelație, 73, 119, 123-126, 129-138, 146-152, 158, 170, 264, 265, 268, 269, 272
- cuantificare, 12, 13, 17, 30
- culoare, 12, 13, 18, 25, 26, 118, 121, 137, 138, 144, 146, 153, 165, 166, 169, 173, 175, 231-240, 250, 252-255, 259, 287
- digital, 10, 11, 77, 95, 102, 229, 232
- distanța
- cosinus, 236, 273-282
 - de autocorelație, 242
 - euclidiană, 20, 239
 - EMD – Earth Mover Distance, 13
- Hafner, 236
- Hamming, 134
 - Hausdorf, 236
 - matriceală Frobenius, 264
 - Manhattan, 13
 - normalizată, 20
 - pătratică, 13, 17
 - vectorială city-block, 264
- energia sonoră, 18
- energie, 9, 107, 110, 111, 112, 113, 230, 248
- entropie, 245
- extragere, 5, 8, 13, 16, 212, 229, 234, 253, 254, 255, 261, 262
- filtru Gauss, 213
- filtru Kalman, 224, 226
- fluorescență, 8, 97, 98, 101, 104, 112
- formă, 26, 41, 71, 79, 135, 136, 157, 192, 194, 232-234, 250, 251, 254, 255, 259, 262, 285, 286, 291-294
- fractal, 249
- Fuzzy-C means, 17
- imaginar, 286, 290
- image, 7, 12-14, 24, 55, 61, 64, 65, 71, 80, 81, 96, 97, 99, 100, 102, 106-109, 112, 116, 119-123, 126, 132, 135, 136, 140-142, 144, 156, 165, 166, 169, 172, 185, 187-189, 214, 217-221, 223, 224, 226, 231, 232, 234-255, 258-263, 268-271, 277, 278, 281, 282, 285-293, 302
- imagini, 2, 3, 5, 7-10, 12-14, 18, 23, 24, 25, 67, 71-76, 81, 82, 84, 91, 98, 104, 106, 111, 112, 114-116, 119-123, 125, 126, 134, 136, 139-141, 143-146, 149, 153, 154, 157, 172, 177, 182, 183, 185, 191, 195, 201, 202, 205-207, 209, 211, 214, 218-220, 224, 226, 227, 229-250, 252-255,

258, 261, 264-266, 268-270, 272,
 281, 282, 285, 287-294, 301, 302
 imagistică, 97, 104, 230, 301
 i-mediat, 284-286, 294
 indexare, 7, 10, 12-14, 26, 233, 282
 informații diegetice, 16
 laser excimer, 108
 luminanța medie, 18
 mass spring, 21
 matrice de similaritate, 17-20, 200
 metrica, 11, 20, 235
 momente, 18, 29, 239, 251, 285, 293
 muchii, 142, 212, 245
 multimedia, 7, 10, 14, 302
 multi-rezoluție, 143, 249
 nanovision, 5, 7-9, 96, 97, 112
 obiect, 13, 23, 33, 38, 41, 51, 55, 56, 58,
 60-62, 69, 71-73, 77, 107, 142, 189,
 191-200, 236, 241, 251, 252, 258-
 260, 287, 288, 291, 293
 oscilator armonic, 21
 percepție, 17, 37, 39, 47, 62, 72, 86,
 257, 285, 287, 291, 292, 293
 pixel DOM, 100
 pixel, 68, 99, 100, 114-123, 125, 126,
 134, 135, 137-139, 143, 144, 146,
 147, 153, 155-157, 160, 164, 169,
 172, 173, 175, 176, 212-214, 231,
 235, 239-241, 245-248, 252, 253
 probabilitate, 13, 22, 137, 163, 167,
 168, 176, 235, 238, 262
 procesare, 2, 143, 230, 239, 240, 254,
 258-261, 301
 proiecție ortografică, 182, 189, 226
 reconstrucție 3D, 5, 8, 182, 302
 reconstrucție afină, 182, 185, 199, 226,
 227
 reconstrucție proiectivă, 182, 203, 205,
 207, 211
 regăsire, 127, 232, 233, 253, 255
 regiuni, 13, 14, 23, 26, 118, 119, 146,
 151, 165-167, 169, 170, 173, 212,
 237, 240, 241, 250-253, 255, 261,
 262
 robot mobil, 182, 211, 226
 scală, 17, 143, 144, 241
 scenă, 27, 28, 30, 57, 115, 116, 129,
 140, 142, 167, 177, 188, 258, 260
 segmentare, 15, 29, 59, 70, 119, 146,
 166, 167, 169, 172-174, 191, 193,
 200-202, 234, 235, 250, 253
 semantic, 14, 76, 232, 255
 similaritate, 8, 11, 18, 19, 20, 123, 132,
 201, 202, 233, 234, 238, 239, 244,
 249, 255, 264, 265, 268, 269
 sistem de referință, 182, 211
 sistem trinocular, 211, 214
 standard, 38, 54, 100, 104, 234, 238,
 244, 246, 265
 statistic, 24, 135
 suprapunere, 32
 textură, 8, 12, 25, 26, 117, 119, 232-
 234, 241-245, 248-255, 258-263,
 270, 271, 273-280, 302
 transformare, 15, 51, 54, 136, 183, 250
 transformata
 – cosinus directă, 254
 – euclidiană, 220, 224
 – wavelet complexă, 265
 – wavelet continuă, 262
 trăsături, 8, 11, 12, 18-20, 22, 23, 25,
 28, 31, 72, 74, 75, 79, 214, 224, 226,
 232, 233, 240, 242, 243, 245, 246-
 255, 257, 260-262
 vector, 11, 16, 18, 20, 30, 121, 144, 161,
 186, 206, 235, 242, 249, 250, 251,
 253, 254
 vision, 6-9, 12, 35, 43, 44, 46-48, 51,
 52, 62, 83, 91-95, 114, 118, 120, 131,
 134, 136, 138, 151, 155, 162, 165,
 177-180, 215, 227, 228, 230, 257,
 260, 284, 302

Termeni abbreviați

ABAplans, 77	LACII, 7
ABC, 31	LCSS, 18
BAS, 104	LIMSI, 62
CER, 26, 27	KBS, 63
CERCO, 60, 87, 90, 258	LM, 63
CENA, 54	LSE, 62
CHO, 106	LPGA, 101
CdTe, 96, 97, 107-113	MARCO, 106
COE, 96	MFCC, 22
CNN, 31	MPEG, 254
DOM, 99-104, 112	MPEG-7, 12, 13, 15, 16, 35, 251
DGPS, 66	MPFB, 74
DTCWT, 257, 262-265, 273, 282	O&M, 63
DTW, 18, 19,	OMS, 39
DMLV, 39	PCB, 102
ELIPSE, 53	POI, 63-65
EMD, 13	PSD, 99-103
EOA, 63	PPD, 101
EXA, 106	RSP, 23, 24
FO, 23, 24	SAMOVA, 11
FR, 23, 24	SIFT, 14, 26
FS, 23, 24	SIG, 63
GQD, 104	SM, 13
GLR, 22	SURF, 14
GUI, 37	ϵ -SVR, 20
HoG, 14	TD-FLI, 112
HMM, 28	TR-CMOS, 77
IJA, 7, 53,	TRM, 29
IRIT, 2, 7-9, 11, 51-54, 61, 80, 84,	VP, 63
87-91	UMPC, 70
INCA, 39	
LAB, 12	

Note despre autori

Silviu Ioan BEJINARIU este absolvent al Facultății de Informatică al Universității „Alexandru Ioan Cuza” din Iași. În prezent lucrează ca cercetător științific în cadrul Institutului de Informatică Teoretică al Academiei Române Filiala Iași și este doctorand în cadrul Școlii de Studii Avansate a Academiei Române. Preocupările sale principale sunt legate de procesarea imaginilor, recunoașterea formelor, fuziunea informațiilor multisenzor, grafică și modelare 2D/3D, sisteme de informații geografice, calcul paralel și distribuit, proiectare și implementare sisteme informatice.

Adrian CIOBANU, cercetător științific, Institutul de Informatică Teoretică al Academiei Române, Filiala Iași din 1990. Absolvent al secției de Electrotehnică a Facultății de Mecanică Universitatea „Transilvania”, Brașov. În cadrul IIT, a dezvoltat cercetări în domenii precum: sisteme cu microcontrolere, rețele neuronale cu aplicații, procesare de imagini. A participat la proiectul european ADIAC, condus de Universitatea din Algarve, Faro, Portugalia în perioada 1998-2001. Reîntors în țară a continuat activitatea în domeniul procesării imaginilor și în prezent are ca interes domeniul regăsirii imaginilor pe bază de conținut, fiind doctorand în cadrul IIT, AR.

Sylvie CHAMBON și-a susținut teza de doctorat în 2005 la IRIT (Institut de Recherche en Informatique de Toulouse) cu tema potrivirii imaginilor color în prezența super-poziționărilor. În 2006 a efectuat un stagiul postdoctoral în imagistică medicală la Télécom ParisTech pe tema înregistrării neliniare multimodale. În perioada 2007-2011 a lucrat ca responsabil de cercetare pentru segmentarea defectelor de suprafață la IFSTTAR (Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux). Din 2011 este conferențiar și își continuă cercetările în domeniul găsirii corespondențelor în imagini și al segmentării, în echipa IRIT/VORTEX (Visual Objects from Reality To Expression).

Mihaela COSTIN este CS III, Academia Română, Iași, doctor inginer în electronică și telecomunicații cu teza „Contribuții la dezvoltarea sistemelor de decizie și aplicații”, a urmat studii post-universitare economice, în psihologie și științele educației. Activitatea didactică (1996-2004, la UAIC, Facultatea de Informatică și UTI, Facultatea Automatică și Calculatoare) și de cercetare în Institutul de Informatică Teoretică AR, http://www.iit.tuiasi.ro/personal/m_costin.html, s-au concretizat în lucrări și proiecte științifice în domeniile: inteligență artificială, științe cognitive, procesări 1D, 2D. Bursier post-doctoral în proiectul POSDRU 56815 „Societatea bazată pe cunoaștere”, a inițiat colaborarea științifică cu IRIT, UPS, Toulouse, concretizată în editarea acestei lucrări.

Alain CROUZIL și-a obținut doctoratul în informatică la Universitatea Paul Sabatier din Toulouse în 1997. Din 1998 este conferențiar la Universitatea Paul Sabatier și membru al echipei de procesare și interpretare a imaginilor (TCI – Traitement et Compréhension d'Images) din cadrul Institutului de Cercetări în Informatică din Toulouse (IRIT). Aria sa de cercetări cuprinde stereoviziunea, „shape from shading”, calibrarea și autocalibrarea camerelor de luat vederi, detecția modificărilor și analiza mișcărilor.

Guillaume GALES și-a primit doctoratul în informatică de la Universitatea din Toulouse în 2011. Teza sa de doctorat a avut ca subiect punerea în corespondență a pixelilor în stereoviziunea binoculară prin propagarea împerecherilor de puncte de interes și sondajul regiunilor. Din 2011 este postdoctorand la Universitatea Națională a Irlandei din Maynooth (NUIM) unde își concentrează cercetările pe aplicații de realitate îmbunătățită pentru aparate mobile care utilizează tehnici de vedere artificială computerizată.

Anca IGNAT este lector dr. la Facultatea de Informatică a Universității „Alexandru Ioan Cuza”, Iași, România, din 1997. A obținut titlul de doctor în matematică în anul 1999 cu o teză referitoare la aproximarea problemelor de control optimal guvernate de ecuații cu derivate parțiale. În perioada 2001-2002 a avut o bursă postdoctorală la Institutul de Matematică al Universității „Karl Franzens” din

Graz, Austria, unde a continuat problemele studiate în teza de doctorat. În ultimii ani, activitatea sa științifică din domeniul procesării de imagini, s-a focalizat pe analiza de textură prin utilizarea transformatei wavelet.

Philippe JOLY, profesor la Universitatea „Paul Sabatier”, Toulouse, Franța, din 2008, doctor din 1996, în domeniul structurării automate a conținutului audiovizual. A lucrat în Universitatea „Pierre et Marie Curie”, Paris 6, membru al grupului MPEG, a contribuit la norma MPEG-7. A participat la înființarea grupului de cercetare „SAMOVA” (structurare, analiză și modelare a conținutului video și audio) în 2002, împreună cu prof. Regine Andre-Obrecht, membru al comitetului director al Rețelei Naționale Franceze RIAM (Rețeaua pentru inovare privind conținutul audiovizual și multimedia), coordonează patru grupuri de cercetare din IRIT în cadrul temei „Analiză și sinteză de semnal și imagine”. Lucrările sale recente se focalizează pe noi strategii de estimare a similarității conținutului video.

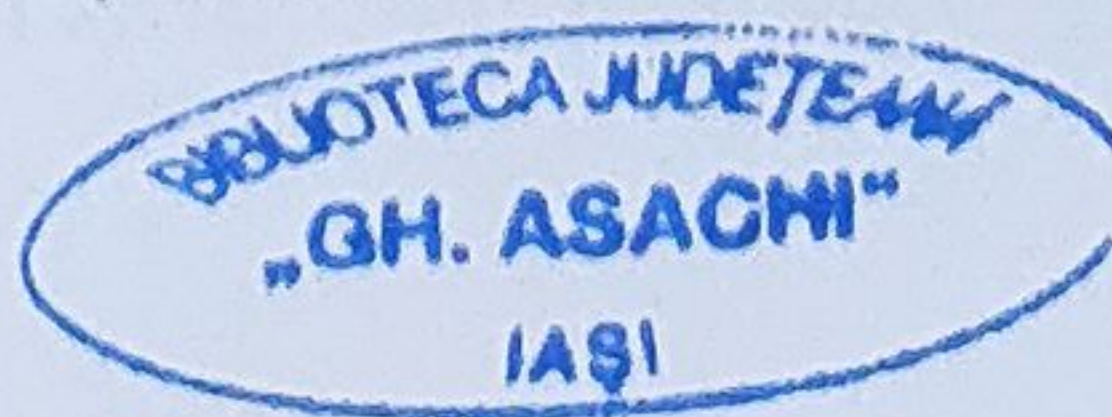
Christophe JOUFFRAIS face parte din Laboratorul ELIPSE, IRIT (UMR 5505, CNRS & Universitatea Toulouse), Toulouse, Franța. Este cercetător CNRS cu studii în domeniul neuroștiințelor cognitive și al tehnologiei de asistare a handicapului. A obținut un doctorat european în anul 2000 în domeniul neuroștiințelor cognitive la Universitatea din Lyon, Franța și în cadrul Universității Fribourg, din Elveția. Domeniile de interes se focalizează pe percepția umană, acțiune și cogniție cu un accent pe tehnologiile de asistare pentru nevăzători. Lista completă de lucrări este disponibilă în site-ul său <http://www.irit.fr/~Christophe.Jouffrais/> din cadrul Institutului de Cercetări în Informatică din Toulouse.

Hidenori MIMURA, director, Research Institute of Electronics, Shizuoka University, Japan, din 2006. Absolvent al Graduate School of Electronic Science and Technology - Shizuoka University, a primit diploma de doctor în 1987. Angajat al societăților Nippon Steel Corporation din 1987, ATR Optic Radio Communications Research Laboratories, din 1994, profesor asociat în Institutul de Cercetări în Comunicații Electrice Univ. Tohoku (1996-2003), este profesor în Institutul de Cercetare în Electronică, Univ. Shizuoka, din 2003, cu domeniul de studiu - dispozitivele imagistice. Visiting-profesor al Univ. Oceanografice - Ocean University, China, Dl. Mimura este și profesor onorific al Institutului de Stat pentru Tehnologie din Sankt-Petersburg.

Simona PESCARU, Ph.D., Chief Vision Scientist, Robotic Vision Technologies, a absolvit Facultatea de Automatizări și Calculatoare, și a susținut doctoratul în cadrul Universității Tehnice „Gh. Asachi”, din Iași cu tema „Tehnici alternative în procesarea semnalului imaginistic”. Cu o experiență de peste 15 ani în domeniul procesării imaginilor și al aplicațiilor computer vision Simona și-a adus aportul în mod esențial în proiectarea de aplicații de recunoaștere și poziționare de obiecte industriale. Aceste aplicații sunt folosite cu succes în fabrici de mașini cu tradiție cum ar fi Ford și GM.

Florin ROTARU a absolvit Facultatea de Automatică și Calculatoare a Universității Politehnica București. Din 1984 este cercetător la Filiala Iași a Institutului de Tehnică de Calcul București, care a devenit în anul 1992 Institutul de Informatică Teoretică al Academiei Române, Filiala Iași. A susținut teza de doctorat cu tema „Contribuții la dezvoltarea metodelor de recunoașterea formelor cu aplicații în analiza imaginilor” (UTI, 1998). Principalele sale preocupări vizează procesarea de imagini, recunoașterea formelor bidimensionale, tehnici de reconstrucție 3D din imagini 2D, dezvoltare de sisteme de prelucrare imagini, recunoaștere de forme, reconstrucție 3D.

Dan Gabriel SÎMBOTIN este doctor în filosofie, specializarea epistemologie și filosofia științei. Cercetător științific în cadrul Institutului de Cercetări Economice și Sociale „Gh. Zane” al Academiei Române – Filiala Iași și Conferențiar la Universitatea „Apollonia” din Iași. Interesat de filosofia științei și de modalitatea în care aceasta atinge momentul de maxim al efervescenței în perioada modernă, a publicat numeroase studii despre filosofia modernă, printre care volumul *De imitatione Dei* (2004). Aria sa de interes cuprinde unitatea științei (a coordonat volumul *Unitatea științei între noutate și tradiție* – Editura Academiei Române, 2011), posibilitățile de cunoaștere, dezvoltarea imaginarului și influența acestuia în construcția imaginii despre lume, în general, și a discursului științific, în special. Este redactor șef al revistei „Logos & Episteme” și redactor al revistei „Symposion”.





Programul Operațional Sectorial pentru Dezvoltarea Resurselor Umane 2007-2013
Proiectul **Societatea Bazată pe Cunoaștere - cercetări, dezbateri, perspective POSDRU /89/1.5/S/56815**
Editori **Academia Română (Filiala Iași) și Institutul European**
Martie 2013

Conținutul acestui material nu reprezintă în mod obligatoriu poziția oficială
a Uniunii Europene sau a Guvernului României

www.proiectsbc.ro | email: proiect.sbc_secretariat@yahoo.com

Volumul a fost editat în cadrul proiectului
Societatea Bazată pe Cunoaștere: cercetări, dezbateri, perspective POSDRU /89/1.5/S/56815,
cofinanțat din Fondul Social European
prin Programul Operațional Sectorial pentru Dezvoltarea Resurselor Umane 2007-2013.
INVESTEȘTE IN OAMENI!

This title was edited within The Knowledge Based Society Project
supported by the Sectoral Operational Programme Human Resources Development (SOP HRD),
financed from the European Social Fund and by the Romanian Government
under the contract number POSDRU /89/1.5/S/56815.
INVEST IN PEOPLE!

The content of these materials does not necessarily represent
the official position of the European Union and the Romanian Government.

ISBN 978-973-611-956-9



www.euroinst.ro



INSTITUTUL EUROPEAN